Aalto University
School of Science
Degree Programme in Computer Science and Engineering

Taufik Akbar Sitompul

Usability and User Experience Evaluation of EUDAT Services

Use Case: Aalto Data Repository

Master's Thesis
Espoo, September 30th, 2016

Supervisor: Assoc. Prof. Keijo Heljanko, Aalto University
Advisors: Assoc. Prof. Antonella De Angeli, University of Trento
Altti Ilari Maarala, M.Sc. (Tech.), Aalto University

Aalto University
School of Science
Degree Programme in Computer Science and Engineering

ABSTRACT OF
MASTER'S THESIS

Author: Taufik Akbar Sitompul

Title:
Usability and User Experience Evaluation of EUDAT Services
Use Case: Aalto Data Repository

Date:	September 30th, 2016	**Pages:**	159
Major:	Service Design and Engineering	**Code:**	T-110
Supervisor:	Assoc. Prof. Keijo Heljanko		
Advisors:	Assoc. Prof. Antonella De Angeli Altti Ilari Maarala, M.Sc. (Tech.)		

The amount of generated research data is growing exponentially. Following this trend, many universities and research institutes require their researchers to plan research data management before commencing a new research project. Research data is a valuable product of research process. Without proper management, the value of research data cannot be utilized properly.

As the largest university in Finland in the field of engineering, Aalto University has defined its research data management policy. The policy aims to make research data management easier and it requires Aalto University to find the most suitable platforms to be used for its research data management. Currently Aalto University has four available options for its research data management platform and EUDAT services are considered as one of the best candidate for Aalto University research data management platform. EUDAT is a pan European project which offers common data services and it is funded by European Union. However, it is still unclear how useful EUDAT services would be for researchers in Aalto University.

This thesis evaluates the current state of usability and user experience of four EUDAT services: B2DROP, B2SHARE, B2FIND, and B2ACCESS. In addition to that, this thesis also evaluates the functionalities that are available on those services. There are two usability evaluation methods that are used to evaluate EUDAT services: feature inspection and usability testing. The results from those two evaluations show that B2DROP and B2FIND have few usability problems and the available functionalities are good enough for researchers in Aalto University. On the other hand, the results show that B2SHARE and B2ACCESS need to be improved significantly before it can be widely used by researchers in Aalto University. This mainly caused by existing usability problems with significant impact that need to be fixed first.

Keywords: usability, user experience, EUDAT, research data, common data services, data management

Language: English

Aalto-yliopisto
Perustieteiden korkeakoulu
Tietotekniikan koulutusohjelma

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä:	Taufik Akbar Sitompul		
Työn nimi:	EUDAT-palvelujen käytettävyyden ja käyttökokemusten arviointi Käyttötapaus: Aalto-yliopiston datavarasto		
Päiväys:	30. syyskuuta 2016	**Sivumäärä:**	159
Pääaine:	Service Design and Engineering	**Koodi:**	T-110
Valvoja:	Professori Keijo Heljanko		
Ohjaajat:	Professori Antonella De Angeli Diplomi-insinööri Altti Ilari Maarala		

Luodun tutkimusdatan määrä kasvaa eksponentiaalisesti. Tätä suuntausta seuraten monet yliopistot ja tutkimuslaitokset vaativat tutkijoitaan suunnittelemaan tutkimusaineiston hallintaa ennen uuden tutkimuksen aloittamista. Tutkimusdata on arvokas tutkimusprosessin tuote. Ilman asianmukaista datan hallintaa tätä arvokasta dataa ei voida kunnolla hyödyntää.

Aalto-yliopisto on Suomen suurimpana insinööritieteiden yliopistona linjannut tutkimusdatapolitiikkansa, jonka tarkoituksena on tehdä tutkimusdatan hallinnasta helpompaa. Tämä myös edellyttää sen, että Aalto-yliopiston on löydettävä sopivimmat alustat tutkimusdatan hallintaan. Tällä hetkellä Aalto-yliopistolla on neljä vaihtoehtoa datan hallinnan alustaksi ja EUDAT-palveluja pidetään yhtenä parhaimpana vaihtoehtoina Aalto-yliopiston tutkimusdatan hallintaan. EUDAT on yleiseurooppalainen projekti, joka tarjoaa yhteisiä datapalveluita ja se on Euroopan Unionin rahoittama. Vielä on kuitenkin epäselvää, kuinka hyödyllisiä EUDAT-palvelut olisivat Aalto-yliopiston tutkijoille.

Tämä tutkielma arvioi neljän eri EUDAT-palvelun tämänhetkistä käytettävyyttä ja niiden käyttökokemuksia: B2DROP, B2SHARE, B2FIND ja B2ACCESS. Lisäksi tutkielma arvioi kyseisten palvelujen toiminnallisuutta. Kahta arviointimenetelmää käytetään arvioimaan EUDAT-palveluja: ominaisuuksien tarkastelu sekä käytettävyyden testaus. Näiden kahden arvioinnin tulokset osoittavat, että B2DROP ja B2FIND -palveluilla on vain vähän käytettävyysongelmia ja niiden toiminnallisuudet ovat tarpeeksi hyviä Aalto-yliopiston tutkijoille. Tulokset osoittavat kuitenkin B2SHARE:n ja B2ACCESS:n tarvitsevan vielä merkittävää kehittämistä ennen kuin Aalto-yliopiston tutkijat voivat käyttää niitä laajalti. Syynä tähän on pääasiassa nykyiset käyttöongelmat, joilla on merkittäviä vaikutuksia ja jotka täytyy ensin korjata.

Asiasanat:	käytettävyys, käyttökokemukset, EUDAT, tutkimusdata, yhteiset datapalvelut, datan hallinta
Kieli:	Englanti

Acknowledgements

This thesis would not have been completed without the contribution of many people. I would like to use this opportunity to express my gratitude to those who have directly or indirectly involved in completion of this master's thesis.

First of all, I would like to thank my supervisor, Assoc. Prof. Keijo Heljanko for giving me an opportunity to work with a big project like EUDAT and supervising this thesis. Working on a big project like EUDAT has been a long and exhausting, but really rewarding experience for me. Special thanks to my two advisors, Assoc. Prof. Antonella De Angeli for her guidance on usability and user experience and Altti Ilari Maarala for proof-reading this thesis for many times.

Thank you very much to thirteen researchers and six students of Aalto University who have participated as test users and pilot users, respectively. Many thanks to developers of EUDAT for their cooperation during these past seven months. I also would like to thank my two colleagues, Farouk Salem and Asad Javed for teaching me to write this thesis in LaTeX.

Last but not least, I would like to thank my family for their unconditional and infinite support during my master studies.

Espoo, September 30th, 2016

Taufik Akbar Sitompul

Abbreviations

EUDAT	European Association for Databases in Education and Training
ACRIS	Aalto Current Research Information System
CC	Creative Commons
MIT	Massachusetts Institute of Technology
CDI	Collaborative Data Infrastructure
API	Application Programming Interface
CKAN	Comprehensive Kerbal Archive Network
CESSDA	Consortium of European Social Science Data Archives
CLARIN	Common Language Resources and Technology Infrastructure
ENES	European Network for Earth System
GUI	Graphical User Interface
HCI	Human-Computer Interaction
ISO	International Organization for Standardization
UEM	Usability Evaluation Method
WebDAV	Web Distributed Authoring and Versioning
SQL	Structured Query Language
LAMP	Linux, Apache, MySQL, and PHP
CSC	Center for Science
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
REST	Representational State Transfer
PID	Persistent Identifier
EPIC	European Persistent Identifier Consortium
XML	Extensible Markup Language
JSON	JavaScript Object Notation
SAML	Security Assertion Markup Language
LDAP	Lightweight Directory Access Protocol
SUS	System Usability Scale

CSUQ	Computer System Usability Questionnaire
QUIS	Questionnaire for User Interaction Satisfaction
SCI	School of Science
ENG	School of Engineering
ARTS	School of Arts, Design, and Architecture
HIIT	Helsinki Institute for Information Technology

Contents

Chapter 1

Introduction

With millions of researchers and scientists pursuing their research across the globe, the amount of generated research data is growing exponentially[1]. Following this trend, many universities and research institutes require their researchers to have a research data management plan before commencing a new research project[2]. Evidence of good practices of research data management has become an expected requirement for granting research funding[3]. Research data management is the procedure that covers research data lifecycle, including creating, storing, delivering, maintaining, archiving, and preserving research data[4].

To have a better understanding of research data management, researchers also need to understand the nature of research data. Boston University stated that "research data is data that is collected, observed, created for purposes of analysis to produce original research results"[5]. Research data can be in physical or digital format and it also comes in various types, such as text, image, audio, video, specimens, samples, and so on. Depending on the research purposes, research data can be classified into four categories, such as observational data, experimental data, simulation data, and derived or compiled data[6]. Different category and type of research data may require different research data management plan.

The advancement of technology revolutionizes the way researchers capture, store, manage, analyse, and visualize research data, thus making re-

[1]http://researchanalytics.thomsonreuters.com/m/pdfs/1003903-1.pdf

[2]https://blogs.ntu.edu.sg/lib-datamanagement/introduction

[3]http://www.dcc.ac.uk/events/workshops/introduction-research-data-management

[4]http://datalib.edina.ac.uk/mantra/datamanagementplans

[5]http://www.bu.edu/datamanagement/background/whatisdata

[6]https://www.nottingham.ac.uk/research/research-data-management/what-is-research-data.aspx

search data management more efficient [17]. Rapid evolution of Internet also makes the world more connected, thus enabling research data to be transferred to other parts of the world easily.

As the largest university in Finland in the field of engineering, Aalto University has defined its research data management policy. In addition to defining the best practices to manage research data lifecycle, Aalto University is also trying to find the most suitable platforms to be used for its research data management.

Currently, Aalto University has four available options for its research data management platforms. The first option is to use research data management platform that has been developed in international collaboration, such as EUDAT services[7]. The second option is to use research data management platform that has been developed by open source community, for example, Dataverse[8], Zenodo[9], Hydra[10], etc. The third option is to use national-level research data platform that has been developed by Finnish Ministry of Education and Culture, for example, Avaa[11], IDA[12], and Etsin[13]. The fourth and the last option is to use ACRIS[14], a research data management platform that is being developed internally by Aalto University based on the Elsevier Pure system[15].

1.1 Aalto University research data management policy

As what have been introduced previously, Aalto University has defined its research data management policy. The objectives of Aalto University research data management policy are to make research data management easier for all researchers within Aalto University and define suitable principles to be used in publishing publicly-funded research data. The policy also describes the guidelines whether researchers should open or close their research data. However, the policy covers research data in digital format only. Research data in physical format is excluded from the policy.

[7] https://www.eudat.eu
[8] http://dataverse.org
[9] https://zenodo.org
[10] https://projecthydra.org
[11] http://avaa.tdata.fi/web/avaa/etusivu
[12] http://openscience.fi/ida
[13] https://etsin.avointiede.fi
[14] acris-test.aalto.fi
[15] https://www.elsevier.com/solutions/pure

Before deciding to open research data, researchers should consider the guidelines for ethical principles and responsible conduct of research that are applicable in Aalto University. Research data should not be opened if the opening of research data might violate privacy, security, safety, funding project agreement, and other legitimate factors. The possibilities to exploit the research results and research data commercially should be taken into consideration before opening research data. In this case, an embargo period can be set in order to let creators of research data take possible advantage of research data before the research data is open for public.

Aalto University research data management policy has defined five principles that should be followed, where applicable, when opening scientific research data. Five principles for open access publishing of research data that have been cited directly from Aalto University research data management policy are as follows [52]:

1. Discoverability

   "The location of research data and the necessary software to access the data is known and they are easily discoverable and identifiable by means of a standard international identification mechanism."

2. Accessibility

   "Research data and the necessary software to access the data shall be easily accessible, along with other necessary information to access the data, such as a mention of the concerned research areas."

   "An embargo period can be agreed upon, to achieve strategic advantage for the creator. The embargo period is expressed in connection with the material."

   "For the research data, a licensing framework has been selected for research and educational use and for commercial exploitation. The license recommended for research data is Creative Commons Attribution 4.0 International (CC BY 4.0)[16] and the waiver recommended for metadata is CC0 1.0 Universal (CC0 1.0) Public Domain Dedication[17]. The license recommended for software is the MIT Licence[18]. Other licenses may be chosen to achieve the strategic goals of the research project. If research data is published using the CC BY 4.0 license, Aalto University and researchers obtain sufficient rights for the re-use of the data, even though the ownership of the research data is not transferred to

[16]https://creativecommons.org/licenses/by/4.0
[17]https://creativecommons.org/about/cc0
[18]https://opensource.org/licenses/MIT

the University. The terms of the CC BY 4.0 license require that the creators of the research data are attributed."

3. Assessibility and intelligibility

   "The research data and the necessary software to access the data shall be assessable for and intelligible to third parties for scientific scrutiny and peer review. Research data shall be published together with related scientific publications for the purposes of peer review. Research data shall be provided in a way that judgements can be made about its reliability and the competence of those who created it."

4. Usability

   "The research data and the necessary software to access the data shall also be usable for purposes beyond the original project. The research data chosen for long-term preservation shall be safely stored and curated to ensure its usability for wider needs than merely a highly specialised expert group. Whenever possible, certified repositories are used for research data preservation."

   "Research data shall be stored together with the minimum metadata and documentation to make it useful. Aalto University does not require the creator of the research data to offer support for further use of the data."

   "When applicable and allowed by previous licenses, the necessary software will also be stored together with the research data."

5. Interoperability

   "The research data and the necessary software shall be produced in a manner that ensures their interoperability with applicable standards and special quality requirements. Interoperability shall be implemented in a way that allows data exchange between researcher groups, higher education institutions and research institutions in different countries. Interoperability will also allow for the re-combination of different datasets from different origins."

The policy also requires Aalto University to have suitable software platforms for opening research data and it should be done in a cost-effective way, where obtainable benefits are greater than resources that have been invested. Lastly, the selected platforms should enable researchers to open their research data easily.

1.2 Introduction to EUDAT services

EUDAT is a pan-European project which covers research data lifecycle as part of Collaborative Data Infrastructure (CDI)[19]. EUDAT offers common data services and it enables European researchers and practitioners from any disciplines to find, access, store, and preserve research data in a secure and trusted environment. Those common data services are distributed across fifteen countries in Europe and the research data is stored in some most powerful super-computing facilities in Europe. EUDAT receives funding from European Union under Horizon 2020 research and innovation programme[20] and it has thirty-five partners across Europe.

Currently, EUDAT offers five data services and one authentication and authorization service that can be used by European researchers and research communities[22]. Figure 1.1 shows six EUDAT services and how those services interact with each other. Following are brief explanations of major EUDAT services:

1. **B2DROP**: A cloud storage service where researchers can store, synchronize, and exchange research data in a secure and trusted environment[23]. B2DROP is designed to be a Dropbox-like service[24] and it is developed on top of OwnCloud[25], an open source software for cloud storage. B2DROP enables its users to define with whom, the duration, and how do they want to exchange their research data. In addition to that, B2DROP also supports automatic desktop synchronization and multiple versioning of research data.

2. **B2SHARE**: A reliable and trusted platform where researchers can store, publish, and share research data in any types or formats[26]. B2SHARE is developed based on Invenio[27], an open source software for digital assets management. B2SHARE guarantees long-term persistence of research data and assigns permanent identifier to every uploaded dataset. In addition to that, B2SHARE enables its users to define access policy of their research data, select community-specific

[19]https://www.eudat.eu/what-eudat
[20]https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020
[21]https://www.eudat.eu/services
[22]https://www.eudat.eu/services
[23]https://www.eudat.eu/services/b2drop
[24]http://alternativeto.net/software/dropbox
[25]https://owncloud.org
[26]https://www.eudat.eu/services/b2share
[27]http://invenio-software.org

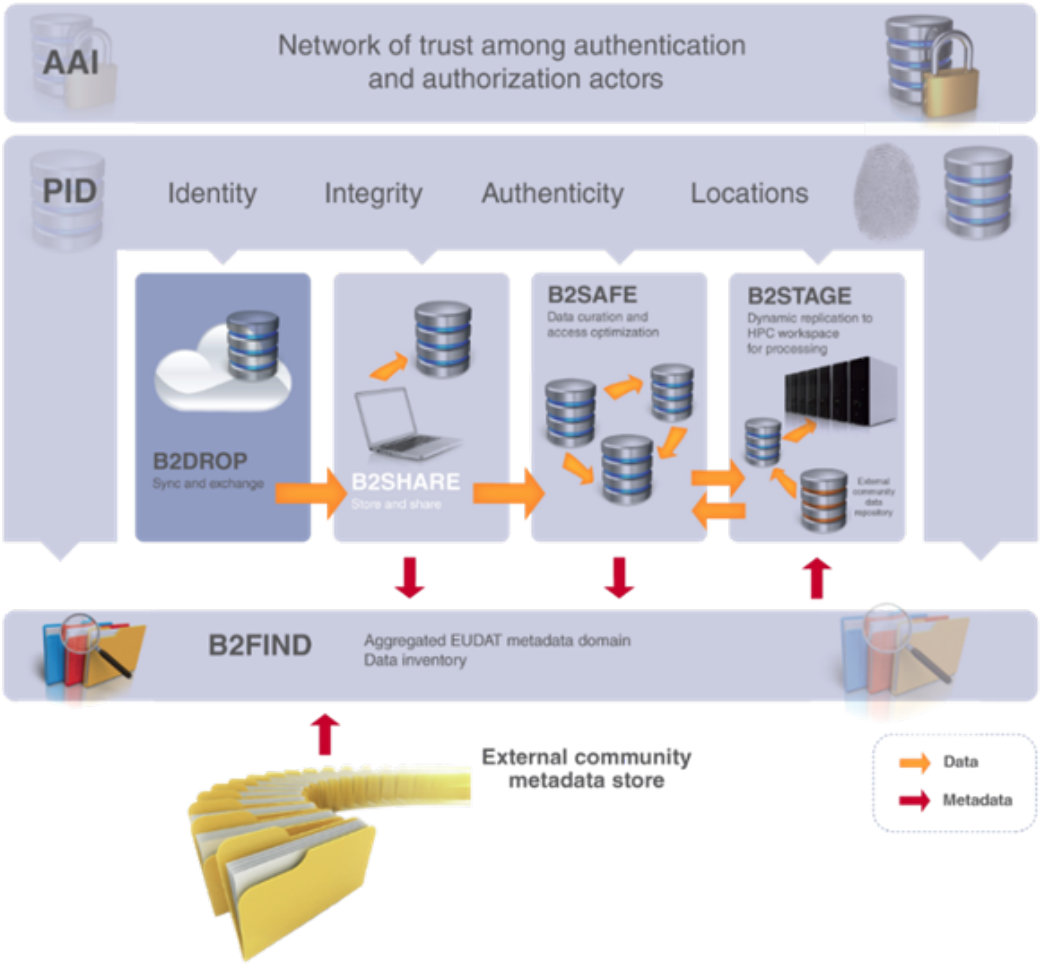

Figure 1.1: EUDAT services[21]

metadata, and upload their research data programmatically by using B2SHARE API.

3. **B2SAFE**: A service which allows research communities to implement research data management policy from various domains in a trusted way[28]. B2SAFE is an ideal solution for research communities who want to virtualize large-scale data resources, avoid data loss during long-term archiving, and improve research data accessibility from various regions.

[28]https://www.eudat.eu/services/b2safe

4. **B2STAGE**: A service which allows researchers or research communities to transfer research data within EUDAT storage facilities or other high-performance computing facilities[29]. B2STAGE is an extension of B2FIND and B2SHARE, where its users can find, store, and preserve research data at the same time. B2STAGE is an ideal solution for researchers or research communities who need storage for large-scale data and high-performance facilities.

5. **B2FIND**: A search engine service where researchers can find and access research data from internal EUDAT services, such as B2SHARE and B2SAFE and also from several external research communities[30]. B2FIND is developed based on CKAN[31], an open source software for data management. Currently, B2FIND can find and access more than 400,000 research datasets from fifteen research communities across Europe[32], such as CESSDA[33], CLARIN[34], ENES[35], etc. In addition to that, B2FIND also allows the researchers to browse research data by using standardized facets and provides overview of available research data.

6. **B2ACCESS**: A secured platform for providing authorization and authorized access to EUDAT services[36]. B2ACCESS allows researchers to log in to EUDAT services by using different methods of authentication, such as organizational IDs, social media IDs, and EUDAT ID. B2ACCESS can be used to create EUDAT ID for researchers who do not have social media IDs or organizational IDs. B2ACCESS is developed based on Unity IDM[37], an open source software for cloud identity and federation management. Currently, B2ACCESS can be used to log in to B2SHARE and B2SAFE only. In the future, researchers will be able to log in to all EUDAT services via B2ACCESS, except B2FIND. B2ACCESS is not required for B2FIND since there is no authentication needed in order to use B2FIND.

As a European-level project, EUDAT is considered as one of the most suitable platforms for implementing parts of Aalto University research data

[29]https://www.eudat.eu/services/b2stage
[30]https://www.eudat.eu/services/b2find
[31]http://ckan.org
[32]http://b2find.eudat.eu/group
[33]http://cessda.net
[34]http://clarin.eu
[35]https://verc.enes.org
[36]https://www.eudat.eu/services/b2access
[37]http://www.unity-idm.eu/site/

management. With thirty-five partners across Europe, EUDAT also offers the best visibility for Aalto University as the source of scientific data. In addition to that, EUDAT is also the most cost-effective solution for Aalto University research data management since it is funded by European Union.

This thesis describes and evaluates B2DROP, B2SHARE, B2FIND and B2ACCESS. Although B2ACCESS does not involve any research data as other EUDAT services, B2ACCESS is included in the evaluation since its important role as the main entrance to access EUDAT services. Currently, Aalto University has not yet considered the usage of B2SAFE and B2STAGE services for its research data management platforms. Thus, B2SAFE and B2STAGE are excluded from the evaluation. More detailed information regarding B2DROP, B2SHARE, B2FIND, and B2ACCESS are presented in Chapter 4.

1.3 Problem statement

As what has been described in the previous section, Aalto University has four available options for its research data management platforms. Currently, Aalto University has also defined functional requirements for its research data management platforms, for example, the system can store metada in addition to the actual data, the system has graphical user interface (GUI), and the system can be integrated with existing user management system [41]. Although some of four available options might meet the functional requirements of Aalto University research data management platforms, it is still unclear whether the selected platforms would be actually useful for researchers in Aalto University. As an example, a system may offer great features that can be used by users to complete their tasks, but those features are useless if the system is too difficult to be used. A system that is difficult to be used may reduce users' productivity or in the worst case, drive users away from using it.

1.4 Thesis scope

The focus of this thesis is to evaluate the current state of usability and user experience of four EUDAT services; B2DROP, B2SHARE, B2FIND, and B2ACCESS. This thesis also evaluates the functionalities that are available in those services. B2SAFE and B2STAGE are excluded from the evaluation since Aalto University has not yet considered those two services for its research data management platforms.

1.5 Thesis structure

This thesis is organized into seven chapters. Chapter 1 introduces the background of growing needs of research data management, Aalto University research data management policy, and EUDAT services. The first chapter also describes the problem statement and the scope of this thesis. Chapter 2 discusses various definitions of usability and user experience and also various usability evaluation methods. Chapter 3 comprehensively describes EUDAT services that are evaluated. Chapter 4 describes the process before, during, and after conducting usability evaluation. Chapter 5 presents the results from the usability evaluation methods and proposes suggestions to improve the usability and user experience of EUDAT services. Chapter 6 describes the conclusion of this thesis.

Chapter 2

Background

2.1 Introduction to usability and user experience

There have been many discussions within human-computer interaction (HCI) community to differentiate usability and user experience [7]. Usability and user experience have become widely accepted terms in academia and industry before definitions and scopes of those two terms have been clearly defined [4]. Usability and user experience have been interpreted differently by different people with various standards [33].

One of the noteworthy efforts to provide clear differences between usability and user experience was made by International Organization for Standardization (ISO). According to ISO 9241-210 [11], usability is "extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use". ISO 9241-210 also mentions the definition of effectiveness, efficiency, and satisfaction in usability. In this case, effectiveness means "accuracy and completeness with which users achieve specified goals", efficiency is "resources expended in relation to the accuracy and completeness with which users achieve goals", and satisfaction means "freedom from discomfort and positive attitudes towards the use of the product".

In comparison to usability, ISO 9241-210 defines user experience as "a person's perceptions and responses that result from the use or anticipated use of a product, system, or service". In addition to that, ISO 9241-210 also points out that "user experience includes all the users' emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviours, and accomplishments that occur before, during and after use".

Based on the definitions from ISO 9241-210, there is an overlapping mean-

ing between usability and user experience. If user experience includes all users' behaviours, it can be assumed that user experience also includes users' effectiveness and efficiency. This analogy seems consistent with some scholars who considered usability as part of user experience [22][28][47]. As an example, Hassenzahl and Tractinsky [22] defined user experience as "a consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.), and the context (or the environment) within which the interaction occurs (e.g. organisational/social setting, meaningfulness of the activity, voluntariness of use, etc.)". Figure 2.1 illustrates differences and relationships between usability and user experience.

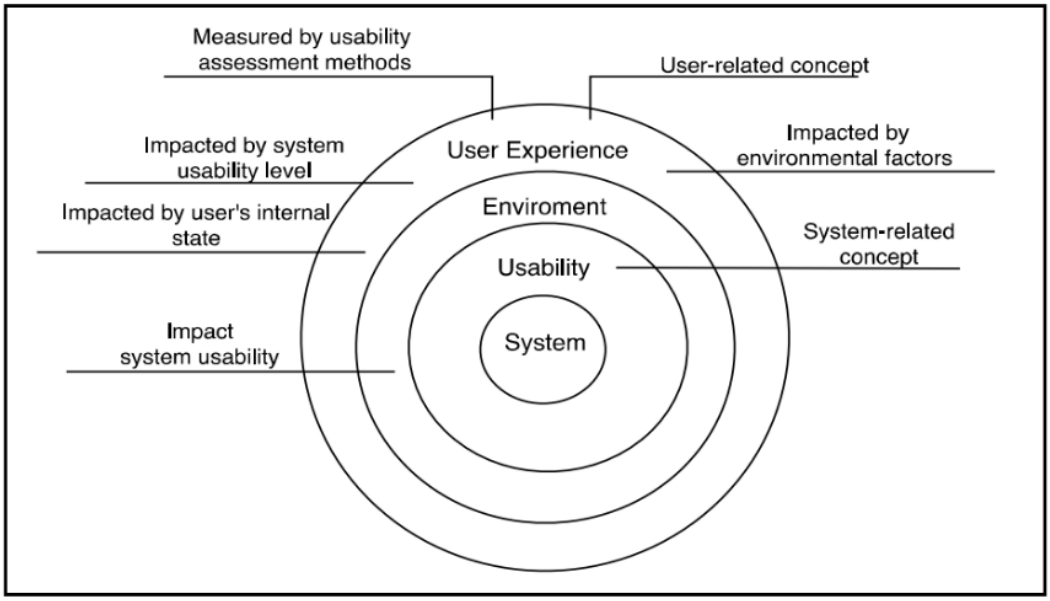

Figure 2.1: Differences and relationships between usability and user experience [33]

Regardless of the terminology, there are some differences between usability and user experience [4]. Usability is usually associated with design of the system and it is used in order to improve human performance when using the system. On the other hand, user experience is usually associated with users' perception and it is mainly used to improve users' satisfaction by achieving pragmatic and hedonic goals. Usability can be assessed by using objective measures, such as required time to complete a task, rate of errors, number of clicks, etc. However, objective measures are not suitable to be used to assess user experience. Nonetheless, user experience can be assessed by using

subjective measures in usability evaluation methods [3][11].

Bringing usability and user experience activities into product development life cycle has been proved to have potential benefits for both end users and vendor companies. This statement is also supported by some scholars within HCI community. Ehrlich and Rohn stated that there are three benefits that will be obtained by vendor companies by bringing usability activities into product development life cycle, such as increased sales, reduced development costs, and reduced costs for providing customer support [14]. From end users' point of view, Mayhew and Mantei [34] stated that end users will benefit from fewer number of errors and increased productivity.

2.2 Usability evaluation methods

Nowadays, there are various usability evaluation methods (UEMs) that can be used. Fernandez et al. [16] defined the usability evaluation method (UEM) as "procedure which is composed of a set of well-defined activities for collecting usage data related to end-user interaction with a software product and/or how the specific properties of this software product contribute to achieving a certain degree of usability". Similar with previous definition, Karat [32] defined UEM as "process for producing a measurement of usability: in evaluation, there is an object being evaluated and a process through which one or more attributes are judged or given a value". One of notable examples of UEMs is the heuristic evaluation introduced by Nielsen [37].

Although many UEMs have been created, there is no shared classification for usability evaluation methods. Different scholars classify usability evaluation methods in different ways [18][46][48]. Several criteria, such as testing environment, user involvement, evaluation goals and data collection methods are used by scholars to classify UEMs [46]. However, the most common way to classify UEMs is to divide it into two categories, namely, usability testing and usability inspection [25]. Generally, usability testing refers to UEMs that evaluate a system by involving test users as the representative of end users, while usability inspection refers to UEMs that assume usability specialists to evaluate the user interface of the system [38].

2.2.1 Usability testing

The first category of UEMs is usability testing. Involving test users of the system is the most fundamental and irreplaceable component in usability testing [37]. The main goal of usability testing is to improve usability of a system by observing how test users actually interact with the system [13].

In addition to that, usability testing can also be used for more specific goals, such as testing system reliability, learnability, and efficiency of the system.

Usability testing consists of a set of predefined tasks that is given to test users by the examiner one at a time. Test users are asked to complete given tasks and the examiner records and observes test users' actions and comments [37]. Most of the time, the test users are asked to do think aloud while completing the given tasks in order to get more information regarding test users' actions and reactions [51]. Thinking aloud means that the test users are required to verbalize their thoughts while trying to complete given tasks [37]. After that, the examiner analyses obtained information, diagnoses usability problems, and suggests solutions to fix usability problems [13].

From the examiner's point of view, the think aloud method offers several advantages [51]. The first advantage is that the think aloud allows the examiner to understand how the test users view the system. The second advantage is that it allows the examiner to detect the areas where users have problems due to interface, design issues, and areas where users have misconceptions. In addition to that, it also allows the examiner to collect large a amount of qualitative data, especially from a small number of test users. However, the think aloud also has disadvantages, especially for test users. The disadvantage of the think aloud is that it might interfere with the way a user would normally behave with the system since test users need to verbalize their actions. The presence of an examiner might also make test users feel uncomfortable. This method can be used throughout the design and development stage, as it relies on the fact that the user is not familiar with the system so that the learning does not mask the problems. It is best used at the early stages of development so that the usability problems could be identified early when fixing them is still cheap [25].

There are some methodological pitfalls in conducting usability testing, such as reliability and validity issues regarding results of usability testing [23]. Reliability means that the obtained results should not change if usability testing with the same methodology was repeated. Validity means that the obtained results should reflect the actual usability issues of the system. In order to obtain reliable and valid results, there are two things that need to be considered carefully. The first thing is that the test users should represent the actual end users and the number of test users should be sufficient [37]. Generally, most scholars assume five people as a sufficient number of test users for usability testing [15][37][39]. However, some studies have shown that more than five test users are required for usability testing since five test users only discover about half of the usability problems [15][59]. The second thing is that the given tasks should represent actual tasks that can be done on the system. Giving unrepresentative or wrong tasks will make results of

usability testing invalid [37].

Usability testing is a long process that requires careful planning before, during, and after conducting usability testing. Well-designed usability testing will save time, money, and effort since the results reflect actual usability issue [1]. On the other hand, poorly-designed usability testing will just be wasting time, money, and effort since the results are not reliable. In general, usability testing can be divided into following tasks:

Develop a test plan

Developing a test plan is the first task that needs to be done before conducting usability testing [12][20][37][50]. The test plan serves as the blueprint since it contains all the required information regarding usability testing that will be conducted. A test plan should address following issues:

- What are the goals of usability testing?
- Where and when will usability testing be conducted?
- How much time is required for each test session?
- How much time and budget is required?
- What are the criteria of the test users?
- How many test users are required and how to recruit them?
- What are the roles of the examiner and who will act as the examiner?
- In what conditions is the examiner allowed to help the test users?
- What kind of tasks that the test users need to perform?
- What kind of test materials (manuals, online guides, questionnaires, etc.) should be provided to the test users?
- How to collect and analyse data from the test users?
- What kind of equipment is required?
- When should the pilot testing be conducted to test the procedure of usability testing?

Choose the examiner

Regardless what kind of UEMs are selected to be used, somebody has to act as the examiner and conduct the evaluation. It is preferable that the examiner knows basics of usability, has previous experience in any of selected UEMs, and has extensive knowledge regarding the system and the user interface that will be evaluated [20][37].

Usability specialists know which UEMs should be used, how to plan and conduct UEMs, and how to collect and analyse obtained data from UEMs that have been conducted [37]. However, usability specialists usually do not have extensive knowledge regarding the domain of the system. On the other hand, the developers of the system have extensive knowledge of the system, but they are not familiar with UEMs. Moreover, it is difficult for the developers to be objective regarding the system that they have developed. Therefore, it is highly recommended that the usability specialists will be the person in charge to conduct usability testing and interact with test users. Developers can still act as the observer and observe the test users from different room. To ensure consistency of usability testing, it is also recommended that the same person conducts all of the test sessions [13].

Find and select test users

The main rule for test users of usability testing is that the selected test users should represent the intended end users of the system [37]. If there are only a few test users, the examiner should only consider test users from the main group of end users. If there are more test users, the examiner should also consider test users from several subgroups of end users. Therefore, the results of usability testing will cover different categories of the intended end users. Friends, family, or colleagues should not be considered as the test users for the usability testing, since the close relationship between the examiner and the test users might produce biased results [20].

Following are several steps that can be followed to select test users [13]:

1. **Develop user profiles**

   User profiles are required to determine the suitability of test users [37]. User profiles can define who are the intended end users of the system, what characteristics do they have, and which characteristics differentiate one subgroup from another subgroup. When the characteristics have been defined, those characteristics should be ranked in order to give priority to more significant characteristics when selecting participants.

2. **Define subgroups of intended end users**

   After the user profiles have been developed, the next step is to define subgroups of intended end users of the system. A subgroup is a collection of the intended end users who share the same defined characteristics in the user profiles. As an example, one subgroup might refer to novice users and another subgroup refers to users who have some previous experience with similar systems [37]. It is also well known that expert users will find less usability problems since they are quite familiar with a similar system. Therefore, novice users are more preferable for usability testing. After the subgroups have been defined, it is required to rank the defined subgroups in order to define which subgroup has the highest priority as test users [13]. If there are many test users participate in usability testing, it is also necessary to define the number of test users from each subgroup [37].

3. **Define required number of test users**

   The number of required test users completely depends on how many subgroups of intended end users should be covered and how much money and time can be spent for usability testing [37]. It is well known within HCI community that five test users are enough for usability testing [15][37][39]. However, it is preferred to have more than five test users since it results in finding more usability problems [15][59].

Design task scenario

The basic rule for planning and designing of task scenario is that the selected tasks should represent actual use of the system [37]. In addition to that, the selected tasks should also cover the most important parts of the system. It is also recommended that developers of the system are involved in selecting test tasks since they have extensive knowledge about the system [20].

Task scenario should be designed to give meaningful purpose for test users. Given instructions should clearly describe what is the goal of each task without telling a test user how to actually complete it [20]. Each task should be given to the test users one at a time and presented in a logical order to ensure that the test users will not feel confused. Moreover, each task should be completely independent. If possible, task scenario should start with the easiest tasks. Therefore, test users can get familiar with the system while keeping them motivated during the usability testing.

Task scenario should be short and written in test users' language instead of using specific terms [50]. In addition to that, there should be time limit for each task. After certain time limit of time has been passed, the examiner

may give more hints or additional instructions in order to help the test users to complete the specific task [20].

Prepare test materials

Test materials are documents that will be used when usability testing is being conducted [50]. Within the HCI community, there are no specific test materials that should be prepared since the required test materials are always different from one test to another. Test materials may have several purposes, such as providing guidelines for the examiner to conduct the test, assistance or help when test users find difficulties to complete given tasks, and tools to collect data from test users and fulfill legal requirements.

Following are the most common test materials that are usually used for usability testing [12][13][50]:

- **Orientation script**: a document which contains verbatim that should be explained to test users before the testing started. Introduction script usually includes several things, for example, the examiner introduces himself, explains the purpose of the test, and describes any forms that should be completed. It is important to note that the orientation script should be kept short and professional. The examiner should always provide the exact same information to each test user.

- **Background questionnaire**: a set of questions to gather test users' background that is usually asked before the testing started. Information from the background questionnaire is useful for the testing team to understand test users' performance and behaviour when the test is being conducted. The questions usually focus on characteristics that may affect test users' performance, such as education, occupation, daily routines, etc.

- **Data collection instruments**: a set of tools that is used to collect information from test users when the test is being conducted. Taking notes during the test is a common practice in usability testing. However, taking notes can be a challenging task, especially if the examiner has to take notes while moderating the test. There is no specific tool that should be used for data collection since it varies from test to test.

- **Legal forms**; a set of documents that should be signed in order to obtain promises and permissions from test users. The most common types of legal forms are non-disclosure, informed consent, and recording permission forms. Regardless of which legal forms will be used, it is

advised to ask the legal department in the company to review them first. It is also strongly recommended to inform the test users in advance if they are required to sign any legal form.

- **Pre-task questionnaire**; a questionnaire that is given to the test users before they started using the system. Unlike the background questionnaire, the pre-task questionnaire is designed to address specific test objectives, such as collecting test users' first impression on the system, determining test users' level of expertise, and classifying test users into specific group of users.

- **Manuals**; a set of documents that contains information regarding on how to operate the system and a description of each feature on the system. Manuals are supposed to be used when test users find difficulties while completing the given tasks. It is strongly recommended to prepare manuals in a professional way with neat text and graphics.

- **Post-task questionnaire**, a questionnaire that is given to test users after they have completed the given tasks. Post-task questionnaire is designed to gather information regarding what test users felt after using the system and discover the system's strength and weaknesses. The questions in post-task questionnaire should be made as efficient and unambiguous as possible to avoid time wasting and misinterpretation. Depending on the test design, post-task questionnaire should be filled during or after the test.

- **Debriefing topics**, a list of general topics that should be discussed before closing a test session. Depending on situation of each test session, the examiner may discuss some or all debriefing topics with test users. If needed, the examiner can also introduce new topics to be discussed during specific test session.

Regardless of test materials that will be used, it is important to prepare test materials well in advance. Make sure there is enough time to include test materials in pilot test and improve them afterwards. This will help the testing team to make structure and organize usability testing in a better way [50].

Conduct pilot test

No usability testing should be conducted before testing what have been planned with some pilot users [12][37]. A person does not have to fulfill

all requirements as a test user, but the person should have no experience with the system that is being tested. In many cases, one or two pilot users are sufficient, although bigger number is preferred.

A pilot test is required to test and improve effectiveness of test procedures and materials before the actual usability testing is conducted. Results from the pilot test are usually not analysed since the objective of pilot test is improving test procedures and not collecting data from users [12]. Results from the pilot test are also useful for the examiner to estimate the duration for each test session and prepare the schedule for the actual usability testing.

Arrange test site and equipment

The test site is an important aspect of usability testing that needs to be considered since the environment may affect performance of test users [2]. Usability testing can be conducted in usability laboratory, test users' site, or even from remote location [21][53][56]. In many cases, usability testing is usually conducted in the usability laboratory [37]. Usability laboratory is preferred since there will be no interruption occurred during the testing and all necessary equipment, such as cameras, computers, and microphones are available already.

Conducting usability testing in usability laboratory is the most convenient option for the examiner [20]. The examiner has full control and there is no need for moving equipment into the test site. However, this option is less convenient for test users since they need to travel to the test site. Test users may also feel uncomfortable or nervous in a different environment, thus affecting their performance during the testing.

On the other hand, conducting usability testing in test users' site is a better option for test users since they may feel more relaxed in a familiar environment. Nonetheless, this option is less convenient for the examiner since the examiner needs to travel and bring the equipment to different places for each test session. Required equipment may vary from site to site. In addition to that, there is no guarantee that no interruption will be occurred during the testing.

Remote usability testing can be an alternative option for both the examiner and the test users. Both the examiner and the test users are not required to travel to the test site and the test users can do the test in any familiar environment for them [53]. However, this option still has some disadvantages. Remote usability testing is not suitable to test any kinds of systems since it is more suitable to test Web-based systems. In addition to that, interruption may occurred if there is problem with the Internet connection.

Regardless of the selected test site for usability testing, it is advised to

simulate the actual environment of intended end users [50]. This would be useful for both the examiner and the test users. Test users may perform like intended end users and the examiner can measure performance of the system in the place where it is supposed to be used, thus resulting in more reliable results.

Conduct usability testing

After the pilot test has been conducted and improvement has been made, real usability testing can be conducted. In general, usability testing can be divided into four stages [12][37][50]:

1. **Preparation**

   Before the test is started, the examiner should ensure that the test site is ready, the equipment is working, test materials are available, and the system that will be tested is functioning properly. Any program or object that is not related to the test should be turned off or moved away in order to avoid distraction during the test.

2. **Introduction**

   When the test users arrive, the examiner should greet them and explain the objective of the test. It is important for the examiner to mention that the test is made in order to evaluate the system and not test user. In addition to that, the examiner should also mention that the test users are free to say anything without being afraid of hurting someone else's feeling.

   The examiner proceeds with introducing test procedures. In case of conducting the test outside the test users' site, the examiner should also explain the computer setup since test users tend to be unfamiliar with them. After that, the examiner can continue with asking test users to fill or sign preliminary documents, such as legal forms, background questionnaire, and pre-task questionnaire. Finally, the examiner can give the task scenario and start the recording.

3. **Run the test**

   When the test has been started, the examiner should observe test users and take notes based on test users' comments and reactions. Generally, the examiner should minimize interaction as much as possible. The examiner should not express personal opinions or show something that may indicate whether test users are doing well or not. However, responding test users' comments with uncommitted sounds like "uh-huh"

is allowed. In addition to that, the examiner should always be aware of his or her voice and body language since they may unintentionally affecting test users.

Generally, the examiner should not help or intervene with test users when they find difficulties, but there are some exceptions for this rule. Firstly, the examiner may help test users if it is clear that test users cannot complete the task, even after referring to available manuals. Secondly, the examiner should intervene if there is equipment or program malfunction. Finally, if the test session takes longer time than what has been expected, the examiner may shorten or even terminate the test session. Before deciding to shorten or terminate a test session, the examiner should encourage test users to complete given tasks first. Although time is being measured during the test, it is important that the examiner should not tell test users how long they have taken the test or what is the time limit for the test.

If there are more than one person who are observing the test, it is required to assign one of them as the official examiner and the others as the observers in order to not confusing test users. The official examiner will be the only one who provides instructions and interact with test users. Other observers should remain quiet and if there are things that should be said, they can do it with passing notes or talking during the break.

4. **Debriefing**

   When the test has ended, the examiner should give post-task questionnaire immediately before having any kind of discussion with test users in order to avoid bias from comments given by the examiner. After that, the examiner may ask several questions to test users based on debriefing topics that have been prepared. The examiner may also ask questions to clarify what has happened during the test. Before closing the test session, the examiner should thank test users and give their compensation, if there are any.

Compile and analyse the results

Usability testing usually generates a huge amount of data. When usability testing has ended, the testing team often has following types of data [13]:

- List of usability problems and system bugs that have been found during the test.

- Quantitative data from performance measurements, such as completion time and number of errors.
- Quantitative data from post-task questionnaire.
- Test users' comments and feedback that are written in the notes.
- Background information from background questionnaire and pre-task questionnaire.
- Recorded video from each test session.

Since the generated amount of data is huge and not all of them can be processed within short period, data analysis can be divided into two different processes with two separate deliverables [50]:

1. **Preliminary analysis**

   A simple analysis that is done to confirm obvious problems quickly, thus designers and developers of the system can start fixing those problems immediately before the final report is ready. Preliminary analysis can be done as soon as the usability testing has ended. The deliverable for preliminary analysis can be a short report, slides, or even verbal presentation.

2. **Comprehensive analysis**

   An extensive analysis that is done to analyse all generated data from usability testing. Comprehensive analysis usually requires two-to-four weeks and the duration may vary depending on the number of test users. Descriptive statistics, such as means, medians, ranges, and standard deviations are often used in comprehensive analysis. The deliverable for comprehensive analysis is considered as the final report and it includes exhaustive description of updated findings from preliminary analysis and also findings that have not been covered earlier.

Report findings and recommendations

Reporting findings and recommendations is the last task that should be done in usability testing [13][20][50]. After all data has been analysed, it is the time to transform the results into findings and recommendations and report them to any stakeholder. Findings and recommendations are two different things, but closely related to each other. Findings are assumptions that are made after looking at results from data analysis, while recommendations are things that should be done based on the findings.

Although findings can be made entirely by the test team, recommendations should not be proposed without involving designers and developers of the system [13]. This is important to be considered whether the test team is part of the company or external consultant, even though the level of collaboration may vary from case to case. Different perspectives are essentials when it comes to proposing correct recommendations. In addition to that, recommendations that are supported by designers and developers of the system are more likely to be implemented afterwards. Recommendations are useless if they are not accepted by people who will implement them, which are designers and developers of the system.

When writing recommendations, there are five systematic approaches that should be considered:

1. **Prioritize solutions that will have widest impact**

   It is important to note that not all solutions can be implemented at the same time. Prioritizing solutions is useful to determine which solution should be implemented immediately and which solution should be implemented later.

2. **Exclude political considerations from the first draft**

   Political considerations should be excluded when writing recommendations, especially for the first draft. This approach is important to maintain objectivity and provide free space when formulating recommendations. Including political considerations too early may reduce the objectivity of proposed recommendations.

3. **Provide short-term and long-term recommendations**

   Proposed recommendations should include short-term and long-term recommendations. Short-term recommendations are changes that can be implemented without adjusting the schedule too much, while long-term recommendations are changes that require long-term planning to be implemented.

4. **Specify which area where more research is required**

   Usability testing often produces new questions that cannot be answered with current results. However, those questions might be answered with having further testing or using different research methods. Therefore, it is recommended for the recommendations to clearly specify which area where more research is required.

5. **Cover all issues thoroughly**

   Proposed recommendations should cover all issues thoroughly. Although solutions with widest impact get more priority, the recommendations should not ignore other issues. Ensure that everything has been addressed properly including problems, findings, and solutions.

2.2.2 Usability inspection

The second category of UEMs is usability inspection. Usability inspection is considered as a non-empirical method for evaluating usability since there are no test users involved in the process [46]. In exchange to that, usability inspection usually includes usability specialists who examine usability-related aspects of user interface by using their experience and knowledge of usability standards [25].

Usability inspection is really useful in practice since it does not require too much time to prepare and conduct the evaluation [40]. This method can be used to discover usability problems easily and quickly during early stage of product development even if the prototype has not been developed yet [8]. However, it does not mean that usability inspection can completely replace empirical methods like usability testing. In fact, results from empirical methods like usability testing are more valid since test users are involved in the process [40].

Like usability testing, there are some methodological pitfalls in usability inspection. Reliability, validity, and thoroughness are the main concerns in conducting usability inspection [49]. Reliability means that the obtained results should be the same if an evaluation with the same condition was repeated. Validity means that an issue that has been identified as usability problem should represent an actual usability problem. Lastly, thoroughness means that the number of usability problems that have been found should cover all usability problem that are actually exist on the system.

Petrie and Power [42] conducted a research to compare usability problems that are found by usability experts and test users. The results show that usability problems found by usability experts tend to be too sensitive since test users did not find some of the usability problems as real problems. In addition to that, usability experts also missed some usability problems that have been found by test users. Nonetheless, usability inspection is still useful method that can be used to find usability problems in addition to usability testing.

Following are the most common usability inspection methods [38]:

Heuristic evaluation

Heuristic evaluation was introduced by Nielsen and Molich in 1990 [40]. Heuristic evaluation is a usability inspection method where group of usability specialists are asked to evaluate a user interface design in order to discover usability problems [35]. Usability specialists assess a user interface design and categorize usability problems based on nine usability principles [36].

Following are nine usability principles that have been introduced by Nielsen and Molich [35][37]:

1. **Simple and natural dialogue**

   The user interface on the system should be made as simple as possible. The user interface should contain required information only and any excessive or irrelevant information should be removed. Moreover, all information should be presented in natural and logical order to avoid confusion.

2. **Speak the user's language**

   Terminology used in the system should be written with words, phrases, and concepts that are familiar to the user. Unless it is mentioned on the requirements, system-oriented terms should be avoided as much as possible. As far as it is possible, information on the system should be provided in the user's native language instead of a foreign language.

3. **Minimize the user's memory load**

   In general, human has limited short-term memory, thus the user should not be required to remember all information on the system. Instructions to operate the system should be made as simple as possible. The user should always be able to find the instructions easily on the system.

4. **Be consistent**

   A specific object on the system should be presented in a same way all the time. It is including the selection of words, format, position, and operation of that particular object. This approach is useful to help the user to recognize the object within short period.

5. **Provide feedback**

   The system should constantly inform the test user regarding what is happening on the system. The feedback should be displayed within a reasonable time and contains meaningful information for the user.

6. **Provide clearly market exits**

   The system should not make the user to feel like being trapped. To give a sense of control for the user, the system should provide an easy way out from the current situation or bring the user back to the previous state. The way out should be visible all the time and can be accessed easily.

7. **Provide shortcuts**

   The system should provide shortcuts for experienced user to perform the task quickly. Shortcuts can also be used to allow the user to navigate on the system quickly.

8. **Provide good error messages**

   Error messages should be defensive, precise, and constructive. Defensive means error messages will blame the system malfunction as the cause of the problem and not the user. Precise means error messages provides clear explanation to describe the cause of the problem. Constructive means error messages should also provide information on how to prevent or solve the cause of the problem.

9. **Error prevention**

   Error prevention is better than having good error messages. The system should be designed in a way where the user will make error messages to appear as little as possible.

In heuristic evaluation, there should be at least two usability specialists who act as the evaluators [30]. Each evaluator conducts a separate evaluation and records all the findings. Findings may include usability problems, existing features that should not be changed, and other concerns regarding the user interface design. After that, the evaluators should discuss the similarities among their findings and proceed with grading usability problems that have been found according to their level of severity.

Number of found usability problems tends to increase if there are more usability specialists who act as evaluators [37]. Two evaluators can discover about 50% of the problems, while three evaluators can discover about 60% of the problems. About fifteen evaluators are required in order to discover 90% of the problems.

There are several advantages of heuristic evaluation. Firstly, heuristic evaluation is very suitable in conditions where time and resources are limited since the results can be obtained about two or three weeks [27][40].

Secondly, heuristic evaluation can be combined with usability testing in two-phase usability evaluation. Heuristic evaluation is conducted earlier in order to identify obvious usability problems [30]. Therefore, test users do not have to struggle the same problems during usability testing.

Heuristic evaluation also has some disadvantages. Firstly, the quality of heuristic evaluation relies on skills and experience of usability specialists. Usability specialists have knowledge and experience in usability methodology, but they may have inadequate domain expertise [30]. Secondly, usability specialists are not actuals users of the system and they may miss some usability problems that can only be found by involving actual users [27][40].

Cognitive walkthrough

Cognitive walkthrough is a usability inspection method that can be used to evaluate how well a user interface design can support exploratory learning [43][45]. Exploratory learning means that the user guesses what should be done next based on any cue that is shown on the system without having to read the manuals or prior training. Cognitive walkthrough is designed to give opportunities for system designers to evaluate their own design during early stages, before any empirical evaluation like usability testing will be conducted.

Before conducting cognitive walkthrough, there are four things that should be defined [57]:

- Users' profiles.
- List of tasks that should be performed on the system.
- List of correct actions that should be performed to complete a specific task.
- Description of how the user interface looks like when users perform something on the system.

Cognitive walkthrough is performed by trying to simulate user's behaviour when they are trying to complete specific task. For each action that needs to be done in order to complete a task, system designers try to describe a story of how users will interact with the system. After that, system designers try to describe what the users are trying to do and how the user interface looks like at this point. Based on what is shown on the user interface, system designers try to analyze whether users know what should be done afterwards. The simulation process continues until the specific task is completed.

There are two main benefits of cognitive walkthrough. Firstly, cognitive walkthrough can be conducted during early stages of system development since a fully-functioning prototype is obligatorily required. Issues related to ease of learning can be identified earlier before any implementation has been made. Cognitive walkthrough can also be conducted within short period since real users of the system are not involved and the results can be obtained quickly.

Cognitive walkthrough also has several drawbacks. Cognitive walkthrough is not suitable for holistic usability evaluation since it focuses on ease of learning only. Therefore, it is not recommended to use cognitive walkthrough as the only usability evaluation method. In addition to that, there might be differences between how system designers simulate and how actual users will behave. For that reason, validity and reliability of cognitive walkthrough highly depends on how similar system designers can simulate mental process of actual users.

Pluralistic walkthrough

Pluralistic walkthrough is a usability inspection method where users, developers, and usability experts meet to perform usability walkthrough [38]. Unlike cognitive walkthrough, pluralistic walktrough does not focus on ease of learning only and representative users and usability experts are also involved as the participants [5]. In pluralistic walkthrough, all participants are asked to pretend as the actual users of the system.

Pluralistic walkthrough is started when the user interface design is presented on the screen to the participants. While pretending as the actual users, the participants are required to write down what they are thinking based on what is shown each time on the screen. After that, the participants should discuss according to what they have written. However, representative users should speak first for each topic during the discussion.

There are benefit and drawback of pluralistic walkthrough [24]. The benefit of pluralistic walkthrough is that the results from this method is more valid and reliable compare to cognitive walkthrough since representative users are involved. The drawback of pluralistic walkthrough is it is not possible to simulate all possible actions since all participants should follow the selected actions that have been chosen by the walkthrough administrator. Therefore, pluralistic walkthrough tends to miss some usability problems if the selected actions are not thorough enough.

Formal usability inspection

Formal usability inspection is an inspection method to review user's potential task performance on a system [29]. Similar to pluralistic walkthrough, formal usability inspection is done by an inspection team that includes system developers, system designers, usability specialists, and possibly users as well [24]. The goal of this inspection is to improve the ease of use of the system by discovering and fixing usability problems from user's point of view [19].

An inspection packet is required in order to do formal usability inspection [29]. An inspection packet is a set of documents that contains information regarding system description, user profiles, and task scenarios. The system description contains a set of screen drawings and explanatory text of it. The user profiles should include, at least, label, education, and experience of the user. For each task scenario, there should be, at least, information regarding user's goal, starting point, and intermediate situation that user may encounter.

Before the inspection started, one of the participant should act as the moderator. The inspection started when an inspector takes the role as a specific user and performs according to specific task scenario. When the inspection team found usability problems, those problems are described in a form. In formal usability inspection, usability problem is specifically defined as an element of the system that makes user feels difficult or unpleasant to complete specific task on the system.

There are six logical steps that should be followed in order to conduct formal usability inspection:

1. **Planning**

   Roles, such as moderator, owner, inspector, and scribe are assigned to each person in the inspection team. After that, the moderator and the owner prepare inspection packets and schedule a kick-off meeting.

2. **Kick-off meeting**

   During the meeting, the moderator and the owner distribute inspection packets to all inspectors. After that, the moderator describes the inspection instruction and things that they are expected from the inspectors.

3. **Preparation**

   Each inspector is working individually in this phase. Each inspector should review and become familiar with the contents of inspection packets. After that, each inspector takes the role as the specific user as

described in the user profile and try to performs the task scenario. Usability problem should be noted whenever the inspector cannot perform the task scenario.

4. **Logging meeting**

   In this phase, the moderator and the inspectors have a meeting to discuss their findings. During the meeting, the scribe records all findings from the inspectors. The records should be displayed by using a projector, thus everyone in the inspection team can see all findings that have been recorded.

5. **Rework**

   The moderator manages a meeting to identify solutions to solve usability problems that have been found. The inspectors identify and propose the solutions. After that, the owner reviews and selects the suitable solutions to be implemented.

6. **Follow-up**

   When the rework has been completed, the moderator gathers all information regarding inspection process from the inspection team. After that, the moderator writes and distributes the inspection report to relevant stakeholders. The inspection is considered complete when the inspection report has been distributed.

Like other usability inspection methods, formal usability inspection also has benefits and drawback. The system developers and designers learn how to evaluate their own system from user's point of view, thus helping them to develop better system in the future [19]. The inspection is also cost-effective since it can be conducted during early stages of system development when the cost to fix problems is still cheap. The main drawback of this method is it requires each person in the inspection team to know the exact things that should be done. Failing to do so will hinder the effectiveness of this method [29].

Feature inspection

Feature inspection is a usability inspection method that focuses on evaluating available features on the system [38]. Feature inspection is suitable to be conducted during development stage [26]. At that stage, the purpose and capability of each function are known. Each feature is analyzed for its

accessibility, understandability, and capability. The examiner checks if a specific feature has following criteria: can be accessed easily, well-named, and functioning as it should be. Moreover, feature inspection can also be used as basis to propose new features on the system [38].

Consistency inspection

Consistency inspection is a usability inspection method to evaluate the consistency of each element of user interface design [38]. Each element of user interface design is reviewed by the user interface expert and system developers [58]. After that, the user interface expert and system developers make decisions regarding what should be modified now and what should be modified later. Issues that can be resolved quickly are implemented right away, while issues that require further discussion are tabled. Consistency inspection can also be used to evaluate whether the prototype or actual system is consistent with the initial design.

Standard inspection

Standard inspection is a usability inspection method where user interface is evaluated according to the standards [38]. The purpose of this method to ensure that the system has the same standards as other systems in the market. Standard inspection is usually conducted by usability experts since it requires good knowledge of usability in order to implement those standards in practice [54].

2.3 Combining usability evaluation methods

Usability evaluation is an important part of user-centered design and it should be conducted possibly in all phases of system lifecycle [46]. Combining UEMs is required since usability is a multi-dimensional concept that should be examined from different angles. The common practice in combining UEMs is using both heuristic evaluation and usability testing [37]. Generally, the heuristic evaluation is conducted first in order to discover and fix as many obvious usability problems as possible. Therefore, test users do not have to struggle with obvious usability problems during usability testing. After changes have been implemented, the usability testing can be conducted in order to find remaining usability problems that cannot be discovered from heuristic evaluation. Several studies shown that heuristic evaluation and usability testing discover relatively different types of usability problems, thus both can be used to complement each other [9][10][31].

The required combination of UEMs is always different from one project to another [46]. Each UEM serves different purposes of evaluation and and it has its own benefits and drawbacks. Several factors, such as the purpose of the evaluation, the type of the system that will be evaluated, and availability of human and capital resources should be considered when selecting UEMs that will be used. As an example, if the time and the budget are limited, the usability inspection methods are more suitable to be conducted [38].

Chapter 3

EUDAT services

3.1 B2FIND

The first EUDAT service that is evaluated for Aalto University research data management platform is B2FIND. B2FIND is designed and developed to be a metadata catalogue for searching research data [55]. B2FIND is able to search research data that is stored internally in B2SHARE and from several external research communities. The primary target users of B2FIND are individual researchers. However, B2FIND is open for everyone and there is no need to have an account in order to use the service.

B2FIND can be accessed by using two ways. The first way is to access B2FIND by using any standard web browser. B2FIND has graphical user interface that can be used to utilize its functionalities. Currently, B2FIND offers several search filters to browse research data, such as location, time, publication year, communities, tags, creator, discipline, language, and publisher. Figure 3.1 shows an example of B2FIND graphical user interface. Those search filters not only allow cross-community search but also specific-community search. Another way to access B2FIND is by using B2FIND API. Researchers can develop their own application and utilize B2FIND API in order to search research data programmatically.

B2FIND is using CKAN as its underlying technology. Before CKAN was selected for B2FIND's underlying technology, there were a few requirements that have been defined by EUDAT. For example, the architecture of B2FIND should be made as modular as possible to allow components modification and avoid any technology lock-in at a later time. After collaborative evaluation conducted by EUDAT, CKAN was selected to be used as B2FIND's underlying technology. CKAN is also an open source software that has been widely

[1] http://b2find.eudat.eu/dataset?q=

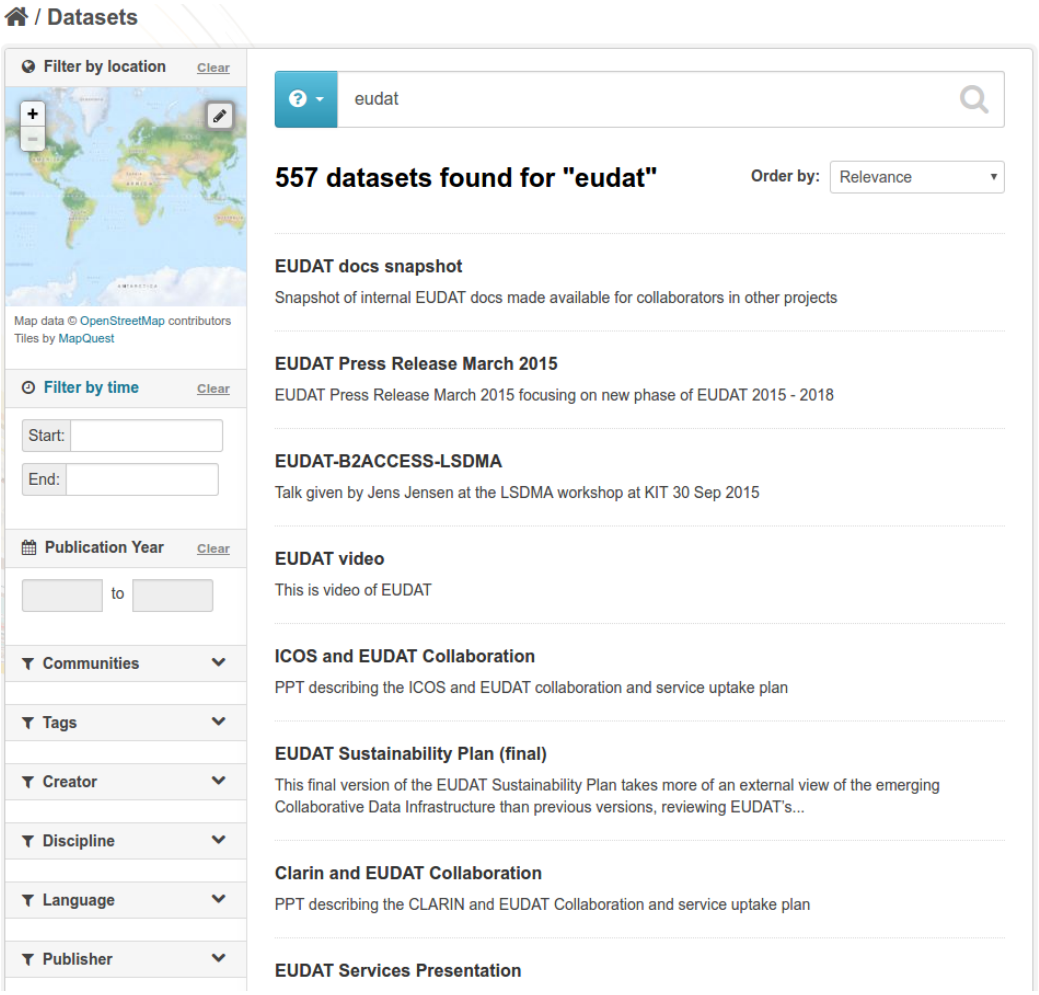

Figure 3.1: B2FIND's graphical user interface[1]

used for not only finding data, but also for publishing and sharing data.

To perform its task, B2FIND utilizes its three modules. The first module is the harvester. The function of the harvester is to collect metadata from provider endpoint by using Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)[2]. The harvester allows B2FIND to search research data that is stored internally in B2SHARE and research data from external research communities. The second module is the mapper. There are several functions of the mapper, such as transforming the format of collected metadata from XML[3] into JSON[4], performing semantic mapping, and performing

[2]https://www.openarchives.org/pmh
[3]http://www.w3schools.com/xml
[4]http://www.json.org

intelligent processing to improve the usability of metadata and reduce the sparse metadata. The third module is the uploader. The function of the uploader is to import metadata records that have been mapped and converted and store them into B2FIND repository. Figure 3.2 illustrates the architecture of B2FIND.

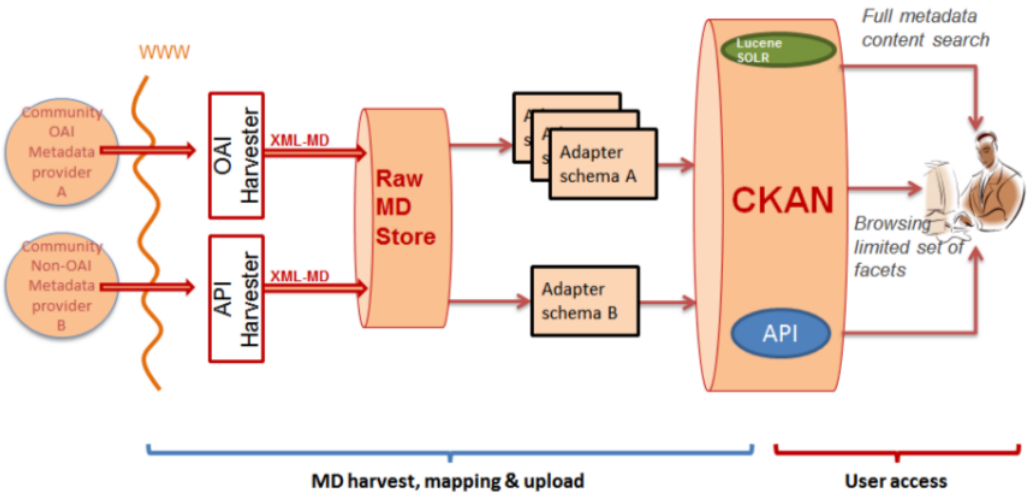

Figure 3.2: Architecture of B2FIND [55]

Currently, there are fifteen research communities and more than 400,000 datasets that can be searched through B2FIND. It is expected that there will be more external research communities joining EUDAT. Therefore, the number of available dataset that can be searched will be increased significantly. The first stable version of B2FIND was released in February 2014 [44]. The development of B2FIND is still continuing until EUDAT project has officially ended in 2018.

3.2 B2DROP

The second EUDAT service that is evaluated for Aalto University research data management platform is B2DROP. B2DROP is a service for storing, synchronizing, and exchanging research data among a team or a specific person [55]. B2DROP allows researchers to store their research data to the cloud storage, exchange their research data privately by using fine-grained access control, and automatically synchronize both local and remote research

data in order to make sure that the uploaded research data is always up-to-date.

The main target users of B2DROP are individual researchers and B2DROP is available free of charge. Currently, the integration of B2DROP and B2SHARE has not been implemented. Therefore, researchers need to create separate new account in order to use B2DROP. Users hold full responsibility to make sure that the uploaded research data in B2DROP does not contain anything which can be considered as illegal according to relevant laws or regulations.

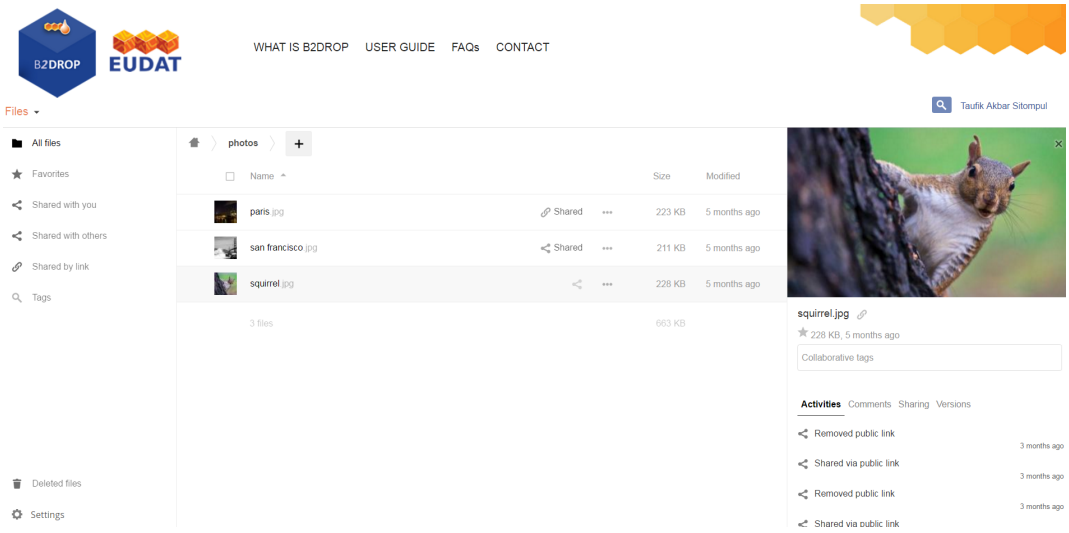

Figure 3.3: B2DROP's graphical user interface[5]

B2DROP can be accessed by using any standard web browser. B2DROP has graphical user interface that researchers can use to access its functionalities. Figure 3.3 shows an example of B2DROP graphical user interface. In addition to that, researchers can use ownCloud client[6] or any WebDAV clients[7] to access their stored research data in B2DROP locally. Installing ownCloud client or any WebDAV client into personal devices also allows automatic synchronization of research data that is stored locally and remotely. Figure 3.4 illustrates high-level architecture of B2DROP's data synchronization and exchange.

B2DROP is using ownCloud as its underlying technology. There were some initial requirements that have been considered before selecting own-

[5]https://b2drop.eudat.eu/index.php/apps/files
[6]https://owncloud.org/install
[7]https://en.wikipedia.org/wiki/Comparison_of_WebDAV_software
[8]https://eudat.eu/services/b2drop

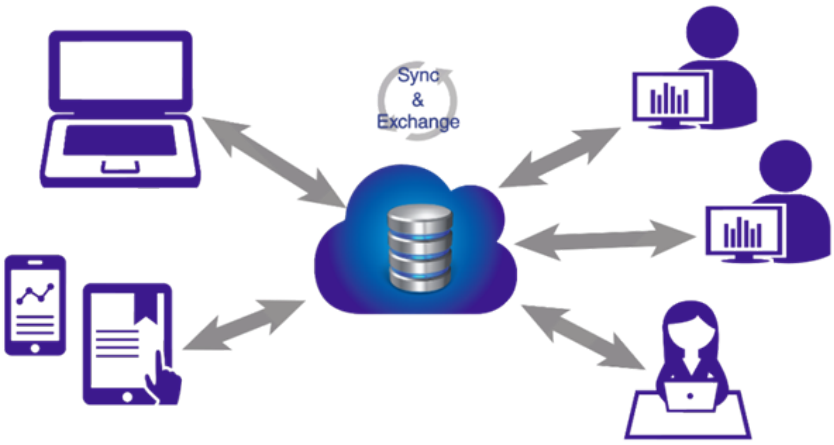

Figure 3.4: B2DROP's data synchronization and exchange[8]

Cloud as the underlying technology of B2DROP, such as compatibility of client-side application with desktop operating systems and mobile operating systems, standard-browser-accessible web interface, and automatic synchronization between local and remote files. In addition to that, there were also additional technical requirements that need to be considered as well, such as file-transfer encryption, scalability, and customizable front-end design. After collaborative evaluation conducted by EUDAT, ownCloud was selected the underlying technology of B2DROP.

The first prototype of B2DROP was available in September 2014 and it was developed on top of ownCloud version 7. B2DROP was developed by using LAMP stack[9], an open source web development framework for deploying dynamic web sites and applications. Currently, the released version of B2DROP is version 2.0. The development of B2DROP is still continuing until EUDAT project has officially ended in 2018.

3.3 B2SHARE

B2SHARE is the third EUDAT service that is evaluated for Aalto University research data management platform. B2SHARE enables researchers to store, publish, and share their research data in any file format [55]. However, B2SHARE is originally designed and developed to support small-to-medium-scale research data or long-tail data, which is often excluded from research data preservation policy in many research institutes. B2SHARE is hosted

[9]https://www.turnkeylinux.org/lampstack

in the data center of CSC in Kajaani, Finland and the research data is also physically stored there [44]. The information security management system of the data center in Kajaani has been certified with ISO/IEC 27001:2005 requirements[10].

The primary target users of B2SHARE are individual researchers. Researchers can log in to B2SHARE by using B2ACCESS. Logging in to B2SHARE is required in order to upload research data to B2SHARE [55]. However, researchers do not need to have an account to browse research data, use metadata, or download publicly available research data. An account is also not needed in order to request restricted research data. Researchers retain the ownership of research data that they have uploaded to B2SHARE and it is the responsibility of the uploader to ensure that the uploaded research data does not contain anything against any applicable laws or regulations. Nonetheless, B2SHARE has the rights to replicate uploaded research data for purpose of long-term preservation [44].

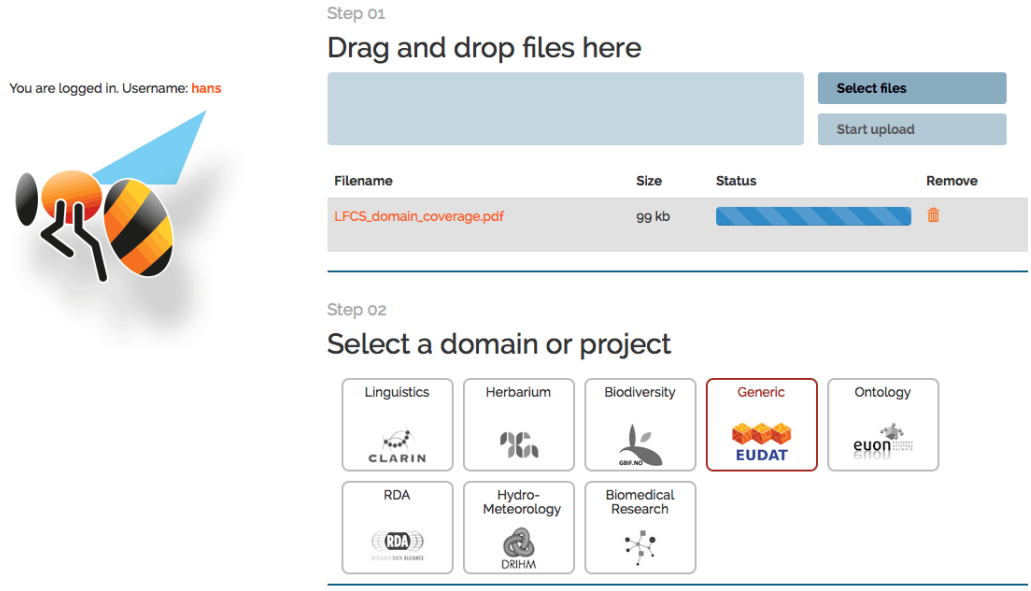

Figure 3.5: B2SHARE's graphical user interface[11]

B2SHARE offers two ways to use and access its service [55]. The first is by using any standard web browser. Researchers can upload their research data through B2SHARE's graphical user interface with web browser. Figure 3.5 shows the interface for uploading research data. B2SHARE also

[10]http://www.iso.org/iso/catalogue_detail?csnumber=42103
[11]https://b2share.eudat.eu/docs/b2share-guide

offers two application programming interfaces (API) that can be used in order to upload large amount of research data automatically. The first API is OAI-PMH API, which can be used for searching and harvesting metadata. For instance, OAI-PMH API in B2SHARE makes stored research data in B2SHARE searchable by using B2FIND. The second API is B2SHARE representational state transfer (REST) API[12], which can be used to upload research data to B2SHARE programmatically.

B2SHARE is using Invenio as its underlying technology and it is developed as an extension of Invenio. In addition to that, B2SHARE takes advantage of Flask[13] for its web framework, Jinja2[14] for its web page templates, SQLAlchemy[15] for its object relational mapper, and MySQL[16] for its database. Moreover, B2SHARE also uses built-in search engine of Invenio to search the metadata. Figure 3.6 illustrates the high-level architecture of B2SHARE.

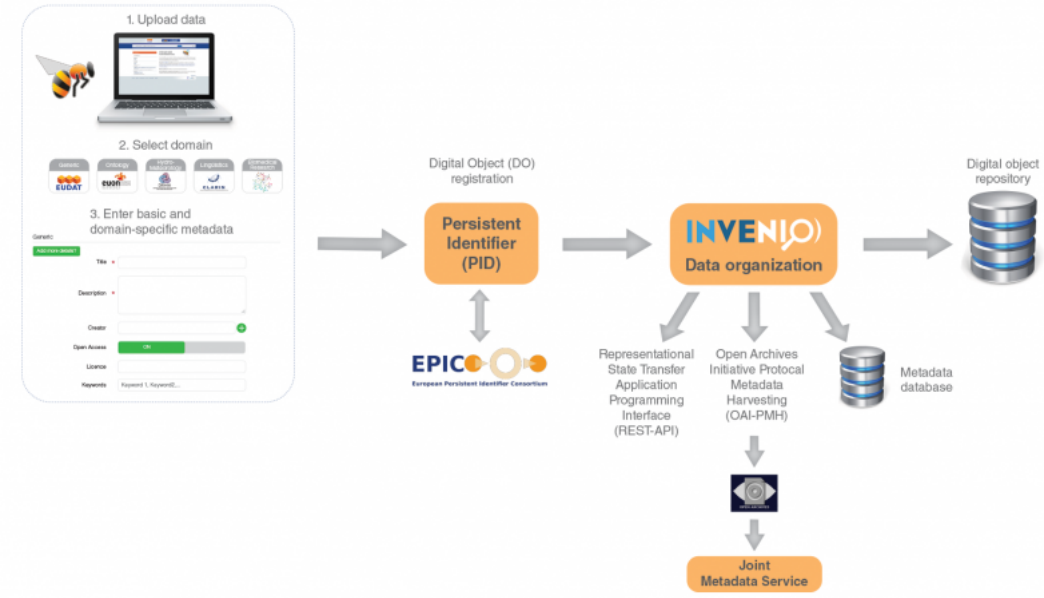

Figure 3.6: High-level architecture of B2SHARE[17]

B2SHARE is using technology from European Persistent Identifier Con-

[12]http://searchsoa.techtarget.com/definition/REST
[13]http://flask.pocoo.org
[14]http://jinja.pocoo.org
[15]http://www.sqlalchemy.org
[16]https://www.mysql.com
[17]https://eudat.eu/services/b2share

sortium (EPIC)[18] to assign persistent identifier (PID) to each uploaded research dataset. PID can be used to trace and find where research data and metadata are actually stored. Metadata is data that provides information about other data and its main purpose is to help users to find relevant information easily[19].

B2SHARE also provides both general and community-specific metadata template that can be used by researchers depending on the needs of their research data. It is possible for research communities to develop their own specific metadata template and make the metadata template available in B2SHARE. Figure 3.7 shows the metadata template of Aalto University that is available in B2SHARE. Currently, there are eleven available metadata templates that can be used in B2SHARE.

B2SHARE is deployed in pre-production environment in January 2014. The first stable version of B2SHARE was released in the end of September 2014 and the current release version of B2SHARE is version 1.0. The development of B2SHARE is still continuing until EUDAT project has officially ended in 2018.

3.4 B2ACCESS

B2ACCESS is the fourth and the last EUDAT service in this evaluation. B2ACCESS is an authentication and authorization platform for EUDAT services. By using B2ACCESS, researchers can log in to EUDAT services by using their existing organizational IDs or social media IDs, such as Google, Microsoft, GitHub, and Facebook accounts. Researchers who do not have organizational IDs or social media IDs can register on B2ACCESS in order to create specific EUDAT IDs. In this case, researchers can use account management tools on B2ACCESS to manage their EUDAT IDs [55]. Figure 3.8 shows an example of B2ACCESS graphical user interface.

B2ACCESS is developed based on Unity IDM[22], an open source software for cloud identity and federation management. Currently, B2ACCESS supports authentication and authorization from SAML[23], OpenID[24], OpenID

[18]http://www.pidconsortium.eu
[19]http://whatis.techtarget.com/definition/metadata
[20]The metadata template was obtained from one of developers of B2SHARE
[21]https://b2access.eudat.eu:8443/home/home
[22]http://www.unity-idm.eu
[23]https://en.wikipedia.org/wiki/Security_Assertion_Markup_Language
[24]http://openid.net/get-an-openid/what-is-openid/

Metadata

Community	`string (required)`	Identifier of the community to which the record has been submitted
Title	`string (required)`	The title of the uploaded resource - a name that indicates the content to be expected.
Description	`string`	A more elaborate description of the resource. Focus on a description of content making it easy for others to find it and to interpret its relevance quickly.
Authors	`array <string>`	The record author(s).
Open Access	`boolean (required)`	Indicate whether the resource is open or access is restricted. In case of restricted access the uploaded files will not be public, however the metadata will be.
Licence	`string`	Specify the license under which this data set is available to the users (e.g. GPL, Apache v2 or Commercial). Please use the License Selector for help and additional information.
Keywords	`array <string>`	A list of keywords that characterize the content.
Contact Email	`[email]`	Contact email information for this record
Discipline	`string`	The scientific discipline linked with the resource.
Embargo Date	`[date-time]`	Date that the embargo will expire.
Contributors	`array <string>`	The list of all other contributors. Mention all persons that were relevant in the creation of the resource.
Resource Type	`array <string>`	The type of the resource.
Alternate identifier	`string`	Any kind of other reference such as a URN, URI or an ISBN number.
Version	`string`	Denote the version of the resource.
Publisher	`string`	The entity responsible for making the resource available, either a person, an organization, or a service.
Language	`string`	The name of the language the document is written in.

Aalto Metadata

This is the blueprint of the metadata block specific for the Aalto community

Project Name	`string`	Project Name
Project URL	`string`	Project URL
Language Code	`string (required)`	This element can be used to add an ISO language code from ISO-639-3 to uniquely identify the language a document is written in
Funding ID	`string`	Funding ID
Owner Organisation	`string`	Owner Organisation
Funder	`string`	Funder
Owner	`string`	Owner

Figure 3.7: Metadata template of Aalto University in B2SHARE[20]

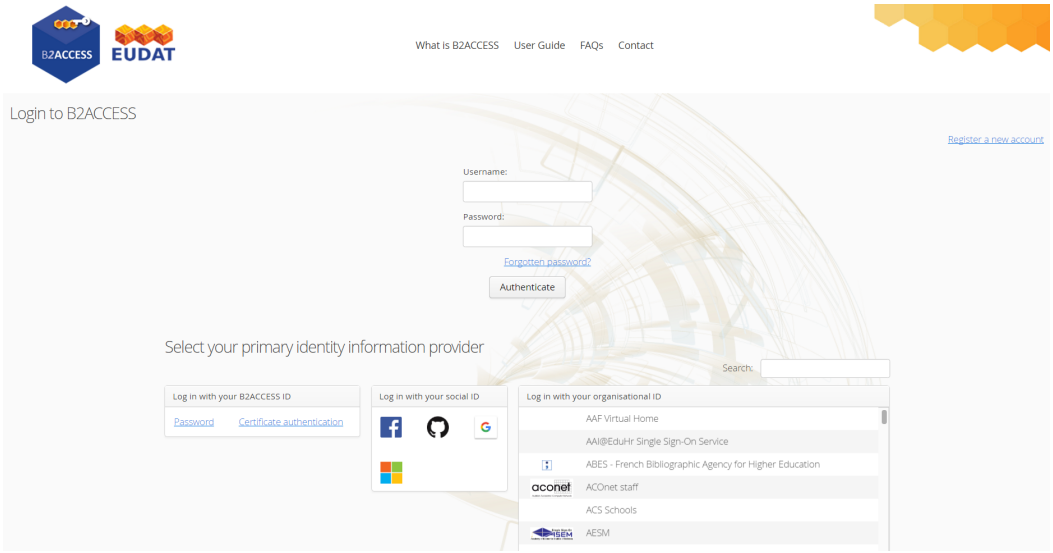

Figure 3.8: B2ACCESS' graphical user interface[21]

Connect[25], X.509[26], and LDAP Authentication[27].

The authentication process starts when B2ACCESS is consuming identity from one of various identity sources. After that, the credentials are used to access different kinds of resources are produced. In B2ACCESS, produced credentials are treated equally regardless the identity sources. Figure 3.9 illustrates the authentication workflow in B2ACCESS.

Currently, B2ACCESS can be used to log in to B2SHARE and B2SAFE only. Researchers and research communities need to create specific accounts for B2DROP and B2STAGE. B2ACCESS is not required in order to use B2FIND since and it can be used by everyone. The integration between B2ACCESS, B2DROP, and B2STAGE is still in progress. It is expected that the researchers can use B2ACCESS in order to use B2DROP and B2STAGE in the future.

[25]http://openid.net/connect/
[26]https://en.wikipedia.org/wiki/X.509
[27]https://docs.oracle.com/javase/jndi/tutorial/ldap/security/ldap.html
[28]https://eudat.eu/services/b2access

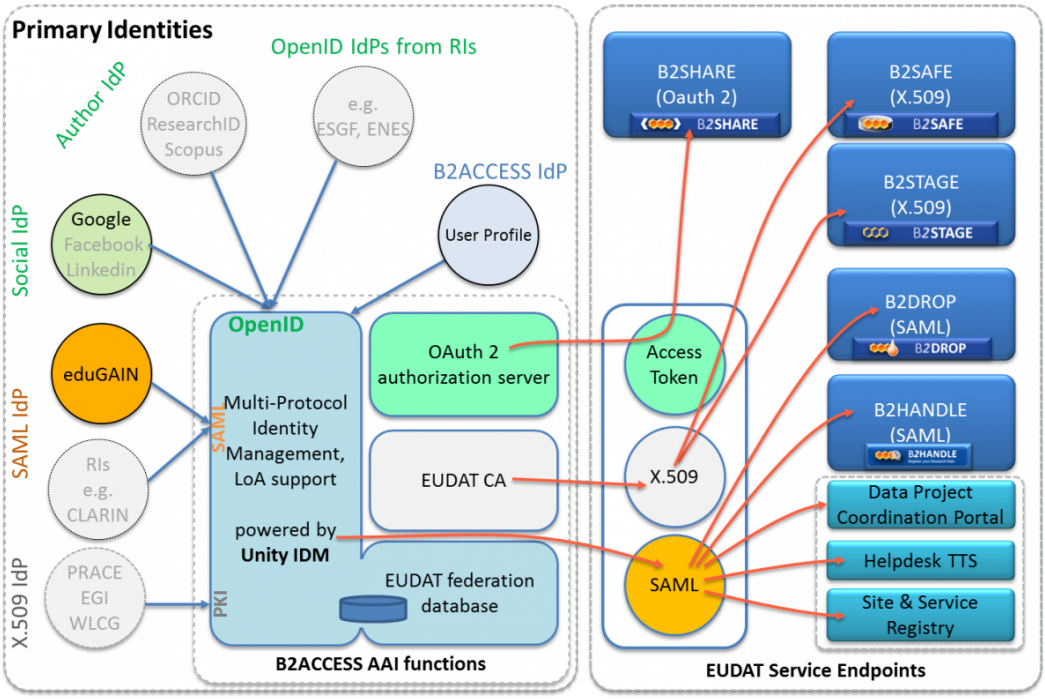

Figure 3.9: Authentication procedure in B2ACCESS[28]

Chapter 4

Conducting usability evaluation

4.1 Selected usability evaluation methods

After considering several factors, feature inspection and usability testing are the selected UEMs to evaluate EUDAT services. Feature inspection is selected since it can be conducted quickly and the results can be obtained within a short period of time. In addition to that, the feature inspection process is also relatively straight-forward. On the other hand, usability testing is selected due to its nature as an empirical method to discover usability issues. It is expected that the results from both UEMs will complement each other.

Heuristic evaluation is not selected since there are no other usability specialists, except the author who conducts the evaluation. Cognitive walkthrough, pluralistic walkthrough, formal usability inspection, and consistency inspection are not selected due to the absence of EUDAT's designers and developers. Moreover, standard inspection is not selected since the author does not have extensive knowledge regarding usability standards.

4.2 Feature inspection

Feature inspection is the first UEM that has been conducted in order to evaluate EUDAT services. The author is the inspector for this evaluation. Conducting feature inspection is quite simple and straight-forward. For each available feature in every EUDAT service, the author inspects its accessibility, understandability, and capability.

The author classifies each usability issues into two categories: the problem and the suggestion. The problem means that a feature does not function as it is supposed to. On the other hand, the suggestion means that a feature is

working properly, but there is something that needs improvement. Moreover, the suggestion is also used to propose new features that would be useful to have on EUDAT services. In total, this UEM discovers sixteen problems and proposes twenty-six suggestions for four EUDAT services (see Table 4.1).

Table 4.1: Numbers of problem found and proposed suggestion from feature inspection

EUDAT service	Problem found	Proposed suggestion
B2FIND	2	1
B2DROP	2	10
B2SHARE	11	15
B2ACCESS	1	0
Total	**16**	**26**

After classifying each usability issues into two categories, the author assigns severity level to each problem and priority level suggestion to each suggestion. There are four levels of severity and priority that can be assigned to each problem and suggestion (see Tables 4.2 and 4.3). Both Table 4.2 and Table 4.3 are adapted from Nielsen's severity ratings for usability problems[1]. See Appendix E for the complete list of problems found and proposed suggestions from this usability evaluation method.

Table 4.2: Severity level of found problems

Severity	Description
Catastrophe	This problem should be fixed before the system can be released.
Major problem	This problem should be given high priority to be fixed.
Minor problem	This problem can be given low priority to be fixed.
Cosmetic problem	This problem does not need to be fixed, unless there is extra time.

Another benefit of this UEM is that it also helps the author to prepare for the usability testing that will be conducted later (see Section 4.3). This

[1]https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems

Table 4.3: Priority level of proposed suggestions

Priority	Description
Urgent	This suggestion should be implemented immediately.
High	This suggestion should be given high priority to be implemented.
Moderate	This suggestion can be implemented later.
Low	This suggestion does not have to be implemented, unless there is extra time.

UEM requires the author to explore all available features on EUDAT services, thus helping the author to determine which features should be tested in the usability testing. This also makes the designing of task scenario for the usability testing easier.

4.3 Usability testing

Usability testing is the second UEM that has been conducted to evaluate EUDAT services. Preparing usability testing is a long process that needs careful planning in order to ensure that the obtained results reflect actual usability issues in EUDAT services. Detailed information for each step in usability testing are described in following sections:

4.3.1 Test plan

A usability test plan dashboard is used to develop the test plan. Usability test plan dashboard is a one-page document that consists nine blocks of relevant information for the test plan[2]. Following are the descriptions that have been filled on each block in the usability test plan dashboard:

1. **Product under test**

   Four EUDAT services; B2DROP, B2SHARE, B2FIND and B2ACCESS will be tested.

2. **Business case**

[2]http://www.userfocus.co.uk/articles/usability_test_plan_dashboard.html

EUDAT is considered as one of the most suitable platforms for implementing Aalto University research data management. However, it is still unclear whether EUDAT services would be really useful for researchers within Aalto University.

3. **Test objectives**

   There are two objectives for the test. The first objective is to evaluate current state of usability and user experience of four EUDAT services. The second objective is to test the capabilities of available functionalities on EUDAT services.

4. **Participants**

   The participant of the usability testing is called the test user. To be qualified as a test user, a person should be a researcher in Aalto University, publish or generate research data, and represent one of the schools in Aalto University. There is no maximum limit for the number of test users.

5. **Test task**

   Each service will be tested with a specific task scenario and each task scenario has different test tasks that need to be completed. However, there is no specific test tasks for B2ACCESS since B2ACCESS can only be used to log in to B2SHARE at the moment. The test tasks for B2ACCESS are included in the test tasks for B2SHARE.

6. **Responsibilites**

   Due to the limitation of human resources, the author is the only person who acts as the examiner and there is no observer involved during the test. The examiner is responsible to develop the test plan, design the task scenario for each EUDAT service, recruit test users, prepare the test materials, conduct the pilot test, conduct all test sessions, analyse the data, and present the results.

7. **Equipment**

   Following equipment is required for each test session: a computer with internet connection, a standard web browser, a video camera, a tripod, and a screen-recording software.

8. **Location and date**

The test is expected to be conducted in any venue within Aalto University. The test will be conducted from April 18th, 2016 until May 13th, 2016.

9. **Procedure**

   The first service that will be tested is B2FIND, followed by B2DROP and B2SHARE. For each service, the test consists several phases. The first phase is introduction, followed by background questionnaire, completing task scenario, filling post-task questionnaires, and debriefing. The whole test is expected to be completed within two hours.

4.3.2 Test users

There are few criteria that a person should meet in order to be qualified as a test user for the usability testing. The first criteria is the test user should be a researcher or at least, a doctoral student at Aalto University. The second criteria is the researcher should produce or publish research data. The last criteria is the test users should come from various disciplines and schools within Aalto University. There is no limitation on gender and age in order to participate as a test user.

In total, thirteen researchers participated as test users for the usability testing. Two test users are researchers in School of Arts, eight tests users are researchers in School of Science, two test users are researchers in School of Engineering, and one test user is a researcher in Helsinki Institute of Information Technology, which is a joint-research institution between Aalto University and University of Helsinki. However, the last criteria of test users is not achieved since no researchers from School of Business are participating as test users within given schedule. Thus, the results of this activity do not represent all schools within Aalto University.

The majority of test users are post-doctoral researchers. There were seven post-doctoral researchers, five doctoral students, one research assistant, and one professor who were participating as test users. Moreover, there was only one researcher who is a friend of the examiner. This is important to note since the results tend to be biased if there are close relationships between the examiner and test users. See Appendix B for the complete profiles of the test users.

4.3.3 Task scenario

Specific task scenario is developed for each service that is being tested. The task scenario contains a list of selected tasks that cover most of the system

functionalities. It is important to note that the selected tasks are designed in a way that test users avoid encountering obvious problems in less-crucial features that have been discovered during feature inspection.

Each service that is being tested has its own task scenario, except for B2ACCESS. Currently, B2ACCESS can only be used to log in to B2SHARE. Therefore, the test tasks for B2ACCESS are included in the task scenario of B2SHARE. In total, there are forty-one tasks that need to be completed by test users within two hours. There are twelve tasks for B2FIND, seventeen tasks for B2DROP, and twelve tasks for both B2SHARE and B2ACCESS. See Appendices C.1, C.2, and C.3 for the complete list of task scenarios.

4.3.4 Test materials

There are several test materials that are used during the test:

1. **Orientation script**

   The orientation script contains information regarding short introduction of EUDAT services, objectives of the test, and procedure of the test. Short introduction of Aalto University research data management policy is also included if the test user is not familiar with it. The orientation script is read before the test is started. See Appendix A.1 for the complete orientation script.

2. **Background questionnaire**

   Background questionnaire contains questions that should be asked in order to understand test users' backgrounds. The questions are asked before a test user starts using a specific EUDAT service. There are two different questions for each EUDAT service. See Appendix A.2 for the list of questions for the background questionnaire.

3. **Manuals**

   The manuals contain necessary information regarding EUDAT services and how to operate them. The manuals were provided in order to help test users when they find difficulties. The manuals were developed by EUDAT and they are publicly available online[3].

4. **Post-task questionnaires**

   Nowadays, there are many types of post-task questionnaires for evaluating usability that are publicly available. For this usability testing,

[3]https://www.eudat.eu/training

three types of questionnaires were used in order to collect as much information as possible. Those three questionnaires are System Usability Scale (SUS)[4], Computer System Usability Questionnaire (CSUQ)[5], and Questionnaire for User Interface Satisfaction (QUIS)[6]

Among those three questionnaires, SUS is the most commonly-used post-task questionnaire for usability evaluation [6]. However, it does not mean that other types of questionnaires cannot be used in addition to SUS. There are few similar questions between SUS, CSUQ, and QUIS. As an example, the question of "The organization of information on the system screens is clear" in CSUQ is similar as the question of "Organization of information" in QUIS. Similar questions among these three questionnaires have been removed in order to avoid redundancy. The words in those three questionnaires are modified as well to suit the nature of the test. As an example, all "product" words in SUS were medified into "system". In total, there are ten questions for SUS, fifteen questions for CSUQ, and eighteen questions for QUIS. See Appendices D.1, D.2, D.3 for the complete list of post-task questionnaires that have been modified.

5. **Debriefing topic**

   Debriefing topic contains default topics that should be discussed with the test users after they have filled the post-task questionnaires. Questions like "what do you feel after using the system?", "What do you like and dislike from the system?", and "Would you like to use the system more frequently in the future?" are the default debriefing topics to be discussed in this usability testing.

4.3.5 Pilot test

Pilot test was conducted two weeks before the beginning of the real test. The test procedure, task scenarios, and test materials are validated with six pilot users. Those pilot users are master and doctoral students in Department of Computer Science, Aalto University.

Results from the pilot test indicated that there were two problems with the post-task questionnaires. The first problem was there are different scaling system between SUS, CSUQ, and QUIS. SUS is using five-point scale, while CSUQ and QUIS are using seven-point scale and nine-point scale respectively,

[4]http://www.measuringu.com/sus.php
[5]http://garyperlman.com/quest/quest.cgi
[6]http://garyperlman.com/quest/quest.cgi?form=QUIS

with an extra option for "not applicable" option. Pilot users also mentioned that it was difficult for them to fill the questionnaires. This is because the scaling system is not meaningful. For example, there is no clear difference whether they should give score seven over eight or vice versa.

The second problem was caused by some statements that are using double negation on the questionnaires. As an example, there is a statement in SUS that mentions "I found the system was unnecessarily complex". Pilot users mentioned that this kind of statement is confusing, especially when the questionnaires should be filled with subjective ratings from strongly agree to strongly disagree.

Based on the results from the pilot test, the post-task questionnaires have been standardized. All questionnaires have the same scaling system with four subjective ratings from strongly disagree to strongly agree and one extra option for "not applicable" option. The changes on scaling system are to prompt test users to pick positive or negative ratings. The statements using double negation were modified into simpler sentences. As an example, a statement of "I found the system was unnecessarily complex" in SUS was modified into "I found that this system was simple". See Appendices D.1, D.2, D.3 for the complete list of modified post-task questionnaires.

4.3.6 Test site and equipment

To ensure that test users are comfortable during the test, the test users were allowed to decide where and when do they want to carry out the test, according to their preferences. It is strongly recommended that the test users carry out the test in their own workplace in order to resemble the situation of actual users of EUDAT services. In addition to that, they were also recommended to use their own computers in the test.

One day before the test, the test users were asked to prepare some real research data that will be used during the test. Using real research data is essential for this usability testing since it can determine whether EUDAT services are suitable or not for researchers within Aalto University. The test users were also asked to install a screen-recording software in their own computers. A screen recording software is required in order to record test users' behavior on the screen. In addition to that, a video camera was also used in the test. The video camera was used to record test users' reactions and comments during the test. Therefore, there is no loss of information after the test, and the videos can be used for further analysis.

In practice, the majority of test users carried out the test in their own workplace. There were four test users who carried out the test in other places, such as conference room and computer laboratory since the conditions in their

workplace were not appropriate.

4.3.7 Real Usability Test

Each test user is given a slot of two hours for carrying out the test. In practice, most of the test users managed to complete the test in less than two hours. The test is also designed to be completed within two hours to ensure that the fatigue level of test users is still in manageable level.

The usability testing is started when the examiner gives introduction about the usability testing and systems that will be tested. After that, the examiner proceeds with asking four background questions to get clear information regarding the test user's background. Then, the test user is given a task scenario to be completed on EUDAT services. While trying to complete the task scenario, the test user is required to verbalize his/her thoughts by doing think aloud. At the same time, the examiner observes and records the test user's actions and reactions.

When the test user has completed a task scenario, the examiner distributes three types of post-questionnaires to be filled by the test user. Once the post-task questionnaires have been filled, the test user is given an opportunity for a short open discussion to describe what he/she is thinking and feeling after using one of EUDAT services. In this stage, the examiner may ask questions based on debriefing topics or other questions that need to be clarified. After that, the test user is given a task scenario to be completed on the next EUDAT service. This process is repeated until the test user has tested B2FIND, B2DROP, and B2SHARE.

To ensure that the test results are not biased, the examiner is expected to minimize interaction with the test user and keep silent during the test. Some exceptions to this rule are if the test user finds some instructions on the task scenarios or questionnaires which are unclear and if the test user still has difficulties to complete a task, even though after referring to the manuals. In these two cases, the examiner is allowed to provide assistance to help the test user.

In practice, not all test users were able to test B2FIND, B2DROP, and B2SHARE. Two out of thirteen test users did not test B2SHARE due to specific reasons. One test user did not test B2SHARE since the generated research data in his research group is highly confidential. The other test user did not manage to test B2SHARE since there was not enough time left. Nevertheless, all of the test users were able to test B2FIND and B2DROP.

4.3.8 Data analysis

Usability testing produces a lot of data that needs to be analysed. All comments from the test users during debriefing periods and while they are completing the test tasks are grouped according to their similarities. After all of the comments have been grouped, the author classifies each comment into two categories: the problem and the suggestion. The problem and the suggestion from the usability testing have the same definitions as the problem and the suggestion from the feature inspection (see Section 4.2). Finally, the author assigns severity level to each problem and priority level to each suggestion (see Tables 4.2 and 4.3). In total, the usability testing discovers twelve problems and proposes sixty-seven suggestions for four EUDAT services (see Table 4.4). See Appendix E for the complete list of problems found and proposed suggestions from the usability testing.

Table 4.4: Numbers of problem found and proposed suggestion from usability testing

EUDAT service	Problem found	Proposed suggestion
B2FIND	2	14
B2DROP	4	28
B2SHARE	4	22
B2ACCESS	2	3
Total	**12**	**67**

In addition to that, several data, such as success rate, completion time, and test users' responses from the post-task questionnaires are analysed quantitatively. The success rate is the percentage of test users who manage to complete a specific test task without help. The completion time is the required duration that a test user needs in order to complete a specific test task. The start time is defined when a test user has finished reading the specific test task, while the end time is defined when the system has completed the operation that is required in order to complete the test task.

Unlike success rate and completion time, test users' responses from the post-task questionnaires cannot be calculated directly since the scaling system has subjective rating. A conversion from subjective rating into numerical value is required in order to allow the author to conduct quantitative analysis. Following are the conversions from subjective rating to numerical value: strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4. Test users' responses for selecting "not applicable" is excluded from

the calculation. The mean and the standard deviation are used to calculate the test users' responses from the post-task questionnaires. See Appendix F for the complete results of the test users' responses from the post-task questionnaires.

Table 4.5: Original and standardized means of test users' responses on three statements in System Usability Scale (SUS)

Original means			
Statements in SUS	B2FIND	B2DROP	B2SHARE & B2ACCESS
Statement no. 4	1.3	1.3	2.1
Statement no. 8	1.6	1.7	2.4
Statement no. 10	1.5	1.4	2
Standardized means			
Statements in SUS	B2FIND	B2DROP	B2SHARE & B2ACCESS
Statement no. 4	3.6	3.6	2.8
Statement no. 8	3.3	3.2	2.5
Statement no. 10	3.4	3.5	3

In addition to that, the one-way repeated measures analysis of variance (ANOVA) is used to test whether there are statistically significant differences on the completion time and the test users' responses from the post-task questionnaires. However, the calculation cannot be done directly since there are two things that should be modified first. Firstly, there are three statements in System Usability Scale (SUS) that are slightly different from other statements from the post-task questionnaires (see Figures F.4, F.8, and F.10). Those three statements are slightly different since lower values of the mean represent more positive responses. Therefore, test users' responses on those three statements need to be reversed in order to standardize the results from the post-task questionnaires before using the one-way repeated measures ANOVA (see Table 4.5). Secondly, not all of the test users have tested B2SHARE, thus making the data size unequal. Therefore, the results from two test user who did not test B2SHARE are excluded for the one-way repeated measures ANOVA. The results from the one-way repeated measures ANOVA are presented in Section 5.1.

Chapter 5

Results

5.1 Overall results

In general, the results from the usability testing are quite satisfying. The results from the one-way repeated measures ANOVA show that the means of overall test users' response are higher than 2.5 (see Table 5.1). From Section 4.3.8, the subjective rating has been converted into numerical value, such as strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4. Therefore, it can be concluded that B2FIND and B2DROP tend to receive positive responses from the test users. On the other hand, B2SHARE and B2ACCESS tend to receive less positive responses from the test users. The significance value that is used in this one-way repeated measures ANOVA is 0.05. The results also show that there is statistically significant difference of test users' responses for each EUDAT service since the p-value is smaller than 0.05 (see Table 5.2).

Table 5.1: Descriptive statistics of one-way repeated measures ANOVA for test users' responses

Test users' responses from the post-task questionnaires	Mean	Standard Deviation	Count
Test users' responses for B2FIND	2.99	0.28	11
Test users' responses for B2DROP	3.12	0.23	11
Test users' responses for B2SHARE & B2ACCESS	2.69	0.29	11

The same analysis is also carried out for the completion time. Table 5.3 shows that, in average, the test users manage to complete the task scenario

Table 5.2: Tests of within-subject effects from the one-way repeated measures ANOVA for test users' responses

Source	Type III Sum of Squares	df	Mean Square	F	p-value
EUDAT services	1.126	2.000	0.563	10.433	0.009

for each EUDAT service in less than 900 seconds or 15 minutes. The results also show that there is statistically significant difference of completion time on each EUDAT service since the p-value is smaller than 0.05 (see Table 5.4). Based on these results, it can be concluded that it is easy to use EUDAT services and researchers can learn to use EUDAT quickly. These assumptions are also identical to the test users' responses from the post-task questionnaires in Figures 5.1 and 5.2.

Table 5.3: Descriptive statistic of one-way repeated measures ANOVA for the completion time

Completion time	Mean	Standard Deviation	Count
Completion time in seconds on B2FIND	411.45	126.21	11
Completion time in seconds on B2DROP	727.81	226.81	11
Completion time in seconds on B2SHARE & B2ACCESS	846.72	198.28	11

Table 5.4: Tests of within-subject effects from the one-way repeated measures ANOVA for the completion time

Source	Type III Sum of Squares	df	Mean Square	F	p-value
EUDAT services	1118538.424	1.960	570590.846	15.894	0.000

Although all test users were able to complete the task scenarios within a short period of time, it does not mean that EUDAT services are flawless and no further improvements are needed. The results from both UEMs shows that there are some functions in EUDAT services that are feasible and some

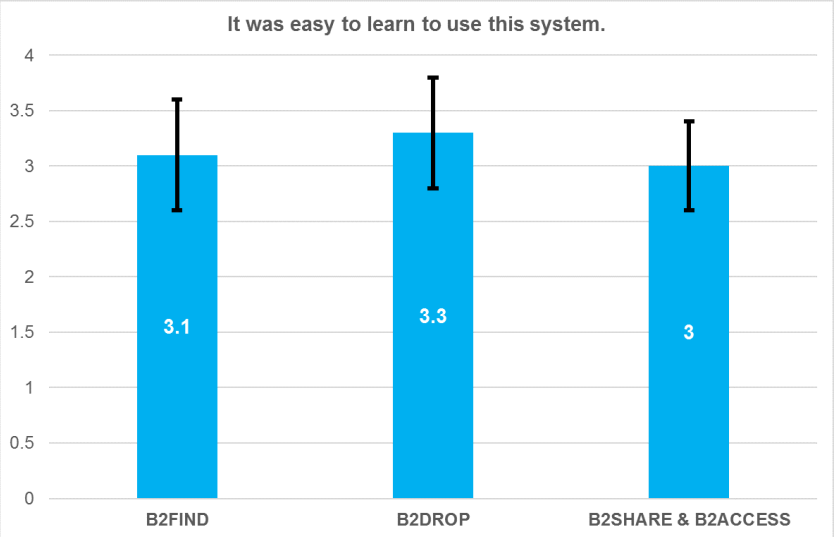

Figure 5.1: The mean scores of test users' responses on statement no. 5 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

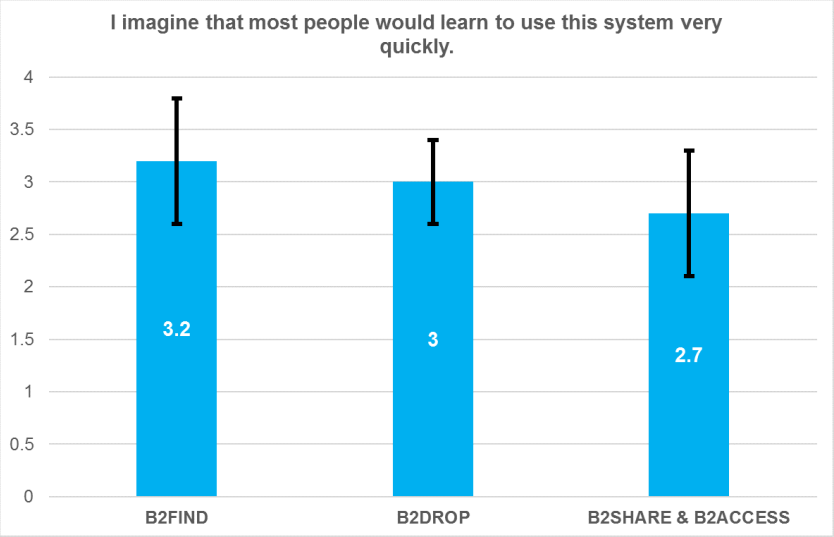

Figure 5.2: The mean scores of test users' responses on statement no. 7 in SUS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

functions that need to be improved significantly. In addition to that, the results from both UEMs also propose some new features that would be useful for the researchers in Aalto University. In total, both UEMs discover twenty-eight problems and proposes ninety-three suggestions. The state of usability and user experience of EUDAT services can be improved significantly by fixing the problems and implementing the suggestions.

Although the number of problems and suggestions are quite large, the results from both UEMs are distinct to each other. There are only three overlapping problems out of twenty-eight problems that have been found from both UEMs. Similar to that, there are only seven overlapping suggestions out of ninety-three suggestions that have been proposed. Therefore, it can also be concluded that both feature inspection and usability testing are complementing each other. The complete results from both feature inspection and usability testing are presented in Appendix E.

Finally, Nielsen's nine usability principles (see Section 2.2.2) are used to describe general findings from both UEMs. Where applicable, test users' responses that are related to the nine usability principles are presented as well. Following are general findings from UEMs that have been categorized based on Nielsen's nine usability principles [37][35]:

5.1.1 Simple and natural dialogue

This principle supposes that the user interface on the system is made as simple as possible. Any excessive or irrelevant information should be removed from the interface. In addition to that, the information should be presented in a natural and logical order to avoid confusion.

Generally, test users agree that B2FIND and B2DROP are simple (see Figure 5.3). On the other hand, test users consider that B2SHARE and B2ACCESS are quite complex. This is mainly caused by the first two tasks in B2SHARE's task scenario which are needed to be completed in B2ACCESS. Currently, it is quite complex to do registration and login process on B2ACCESS. The issue regarding the complexity on B2ACCESS is discussed in more detail in Section 5.5.

5.1.2 Speak the user's language

This principle supposes that the used terminologies are familiar to the user. Using system-oriented terminologies should be avoided as much as possible. Generally, test users agree that the terminologies that are used in EUDAT services are always related to the task (see Figure 5.4).

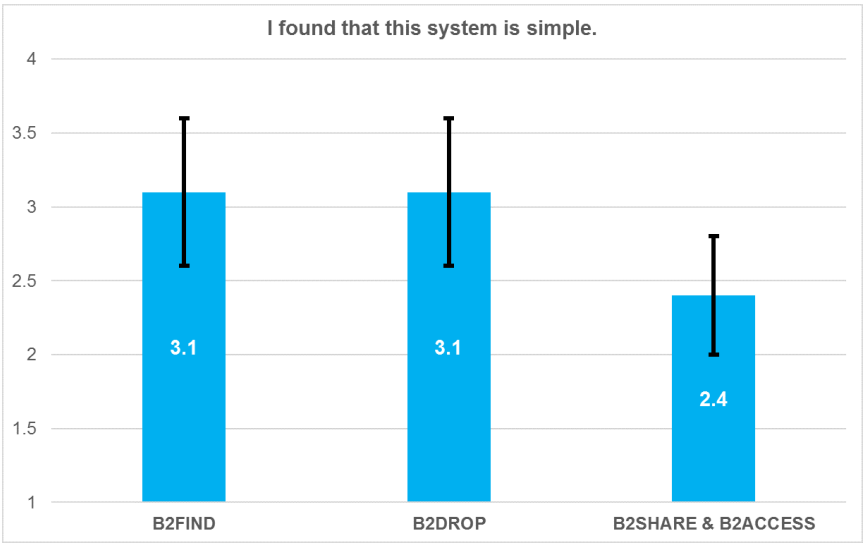

Figure 5.3: The mean scores of test users' responses on statement no. 2 in SUS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

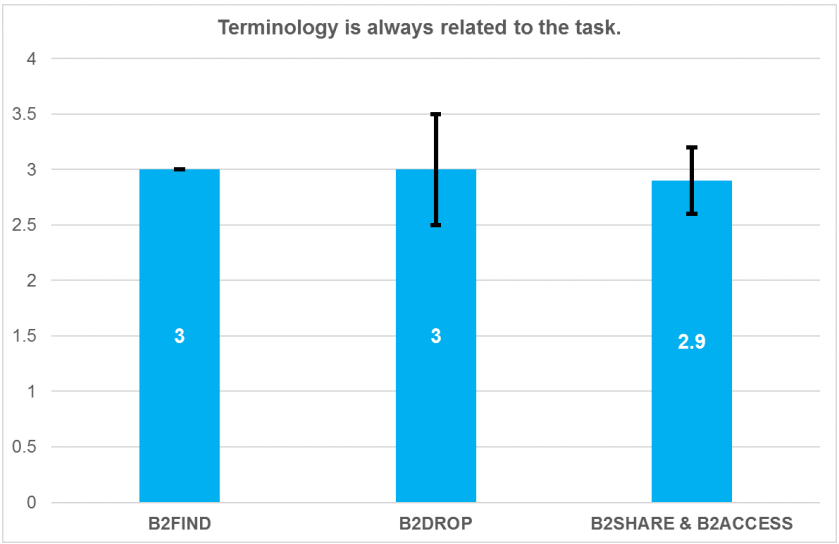

Figure 5.4: The mean scores of test users' responses on statement no. 7 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

There are two minor issues in B2DROP and B2SHARE regarding this usability principle. The username is written instead of email on the login form of B2DROP, although email is the one that is required to log in to

B2DROP. With B2SHARE, there are some test users who are not familiar with some terminologies, such as domain, deposit, embargo, and license.

5.1.3 Minimize the user's memory load

This principle supposes that the user does not have to remember all information on the system. Test users agree that it is easy to remember names and use of commands in EUDAT services (see Figure 5.5). The issues that are related to this usability principle can be found in B2FIND and B2ACCESS. As for B2FIND, it is quite difficult to do filter by time on B2FIND since there are uncommon operations and many steps that need to be taken into account. Further discussion regarding this issue is included in Section 5.2. As for B2ACCESS, this issue is also related to the complexity of registration and login processes. This issue is included in Section 5.5.

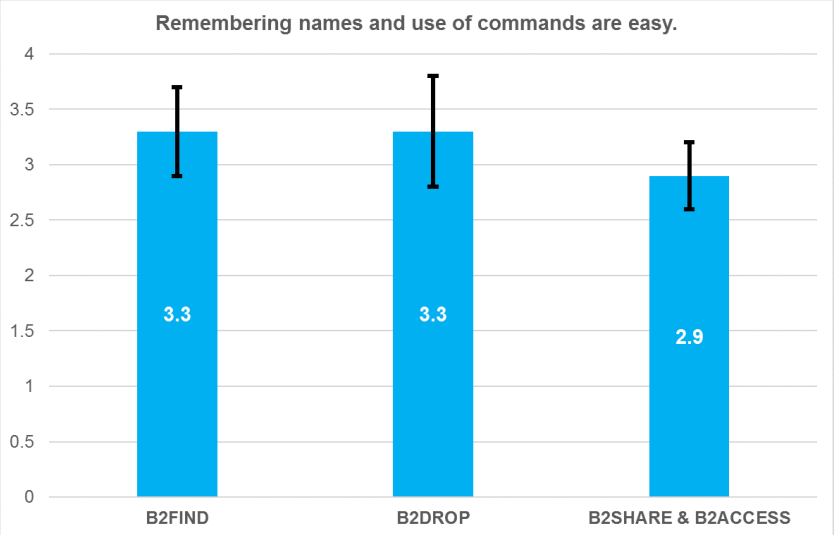

Figure 5.5: The mean scores of test users' responses on statement no. 12 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

5.1.4 Be consistent

This principle supposes that a specific object on the system should be presented in the same way all the time. It can be text, format, position, or operation of that particular object. In general, test users agree that all EUDAT services are consistent (see Figure 5.6). However, some notable issues

regarding this usability principle can be found in B2FIND. There are some search filters in B2FIND that are applied directly and some which are not. In addition to that, there are some search filters from which the user can select one option only and there are some search filters from which the user can select multiple options. This issue is discussed in more detail in Section 5.2.

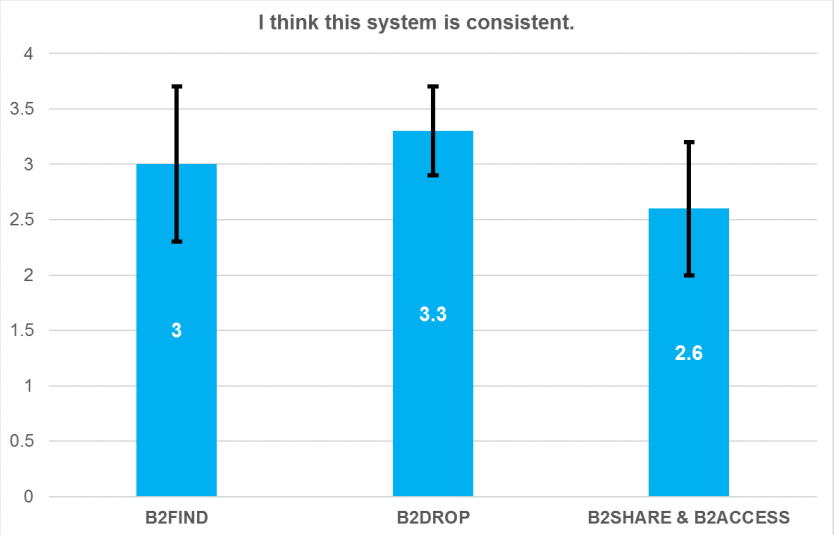

Figure 5.6: The mean scores of test users' responses on statement no. 6 in SUS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

5.1.5 Provide feedback

This principle supposes that the system should always inform the user about what is happening on the system. In addition to that, the feedback should contain meaningful information for the user. Test users agree that B2FIND, B2SHARE, and B2ACCESS always inform about their progresses (see Figure 5.7). However, a notable issue regarding this usability principle can be found in B2DROP since it does not always inform the user about its progress. As an example, when a file is being deleted, the file will simply disappear and the system does not inform the user that the specific file has been deleted.

5.1.6 Provide clearly marked exits

This principle supposes that the system should provide a way out from current situation or bring the user back to the previous state. In addition to

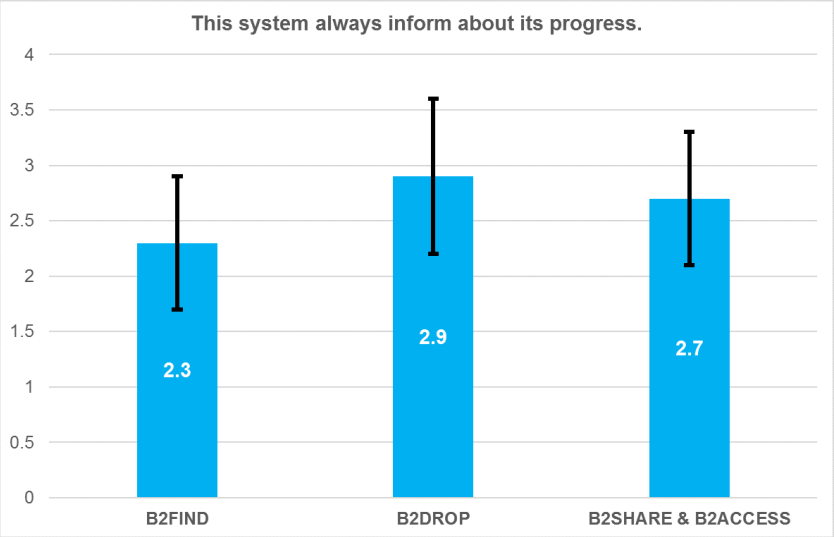

Figure 5.7: The mean scores of test users' responses on statement no. 10 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

that, the way out should be visible and accessible all the time. Some good examples of the way out are by providing undo and cancel button on every system dialogue. Generally, test users notice that they can recover easily and quickly after making a mistake on EUDAT services (see Figure 5.8). Regardless test users' responses, a notable issue regarding this usability principle can be found in B2DROP. Currently, there is no undo feature in B2DROP and operations that have been done cannot be undone.

5.1.7 Provide shortcuts

This principle supposes that the system provides shortcuts for experienced users to navigate and perform the tasks quickly and easily. After conducting the feature inspection, it has been found that there are no shortcuts in EUDAT services. All operations should be done by interacting with the elements in the user interface. In addition to the user interface, each EUDAT service has its own client application or API that can be used to process large amount of data automatically. B2DROP can be accessed by using OwnCloud's client application, while B2FIND and B2SHARE can be accessed by using their own APIs. However, those are not considered as shortcuts since the way to use EUDAT services via user interface and client application or with APIs are completely different.

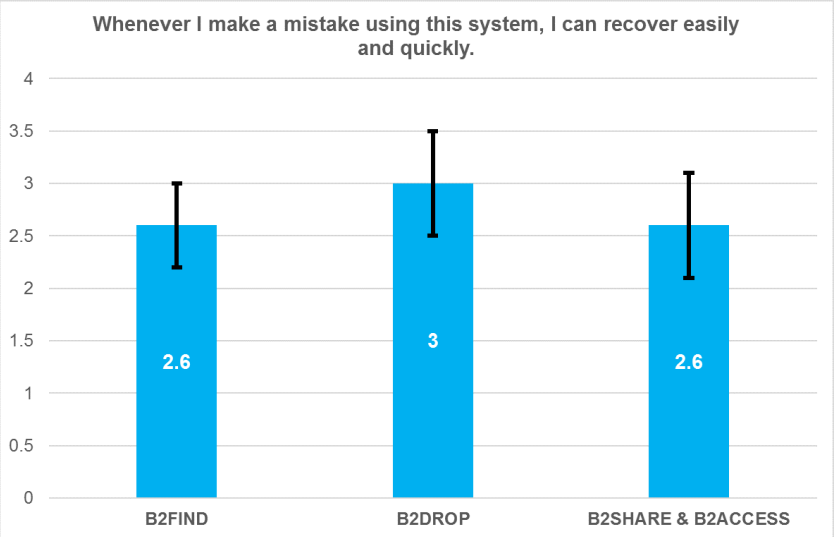

Figure 5.8: The mean scores of test users' responses on statement no. 8 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

5.1.8 Provide good error messages

This principle supposes that the error message describes the cause of the error clearly and informs the user how to prevent or solve the error. Generally, the test users agree that all EUDAT services do not provide well-formed error messages (see Figure 5.9). In case of B2FIND, the system displays internal server error without any additional information when an error has been occurred. In case of B2DROP, the system does not show any error message, except when the user has put wrong email or password. As for B2ACCESS, the error messages display the cause of the errors, but do not explain how to solve the errors (see Figures 5.26 and 5.27).

5.1.9 Error prevention

This principle supposes that the system should be designed in a way the user will produce as little errors as possible. One of good examples to prevent an error from being occurred is by providing confirmation dialogue on the system. The system will ask for user's approval before performing the operation, thus decreasing the possibility of unintentional error being occurred.

The results from both UEMs show that notable issues regarding this usability principle can be found in B2DROP. Currently, the sharing will be applied directly as soon as the user has selected another user to share with.

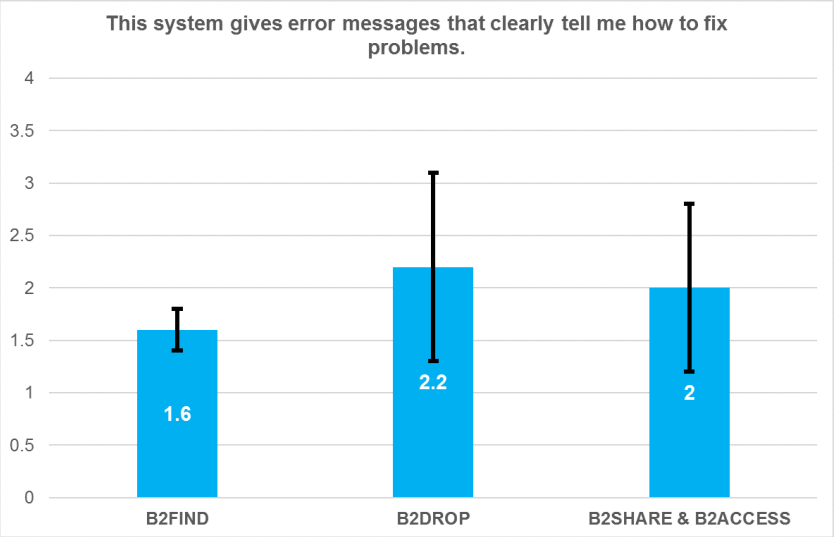

Figure 5.9: The mean scores of test users' responses on statement no. 7 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

The sharing is automatically applied before the user has chosen the access control. In addition to that, files or folders will be permanently deleted from the trash bin as soon as the user clicks the delete button. There should be a confirmation dialogue in order to prevent sharing with wrong person or deleting files or folders that are not supposed to be deleted.

The only issue regarding this usability principle can be found on B2SHARE when a user has uploaded a dataset with open access. The user cannot edit uploaded dataset that has open access. Moreover, the system does not ask for confirmation before the user is uploading his/her dataset. However, the uploaded dataset is editable if it has restricted access.

5.2 Results for B2FIND

There are twelve tasks that need to be completed in B2FIND's task scenario (see Appendix C.1). From twelve tasks, Task 03 and Task 07 are the only tasks where not all test users were able to complete the tasks without help (see Figure 5.10). The results from both UEMs show that the filter by time on B2FIND is very difficult to do and all of the test users fail to complete Task 03 without help. In addition to that, Figure 5.11 also shows that the Task 03 requires the longest average time to be completed.

Inability to complete Task 03 is caused by uncommon and not straight-

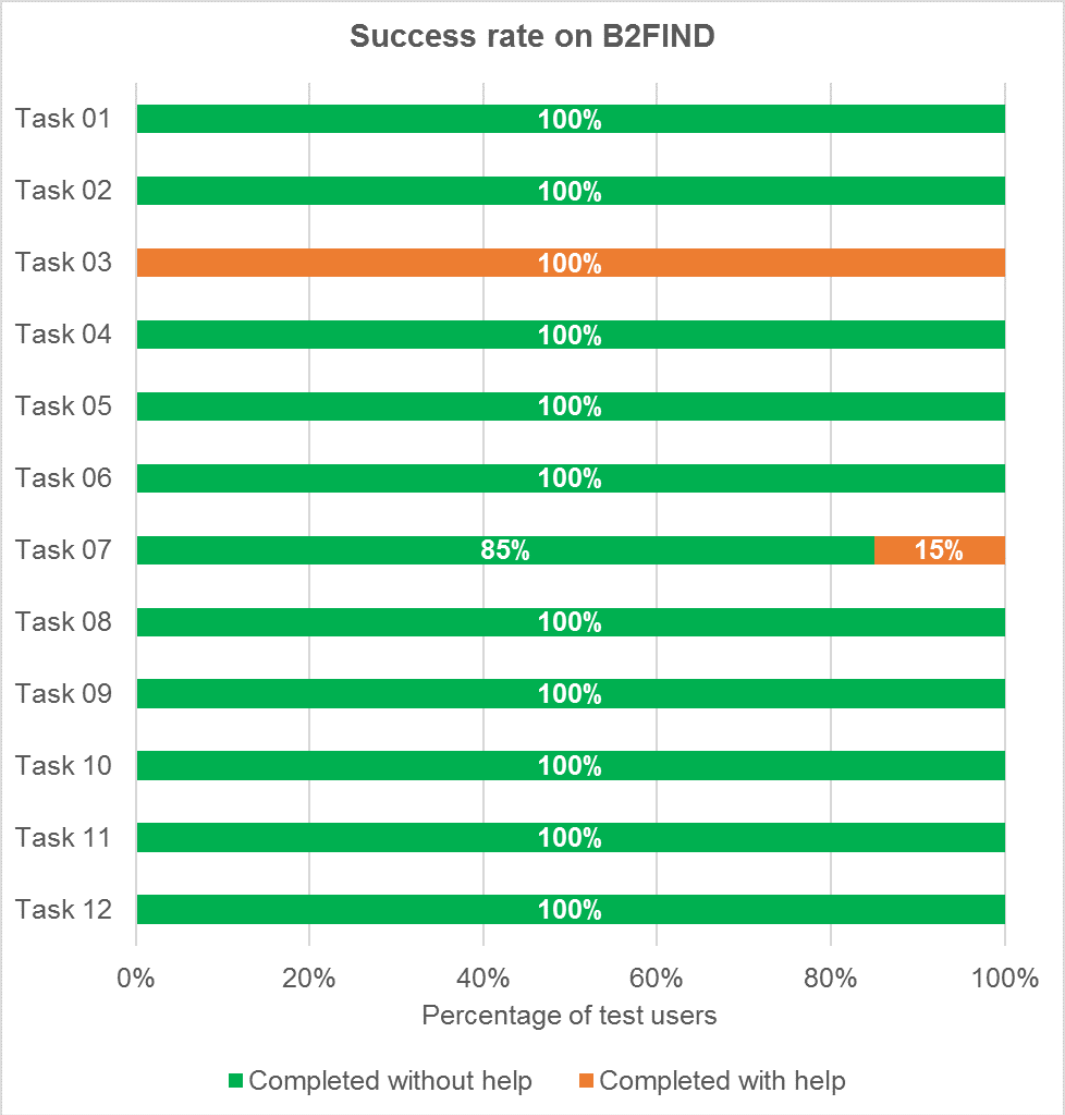

Figure 5.10: Success rate on B2FIND

forward operations that need to be done in order to do filter by time. A system dialogue appears when the user has clicked filter by time. After that, the user needs to select a base period by dragging the mouse on the histogram below and zoom in the period by dragging the mouse on the histogram above (see Figure 5.12). The user needs to select start and end period by holding down Ctrl-key (on Windows and Linux) or Cmd-key (on Mac) and clicks appropriate points on the histogram. After that, the user needs to click the apply button (see Figure 5.13). However, it does not end there. After clicking

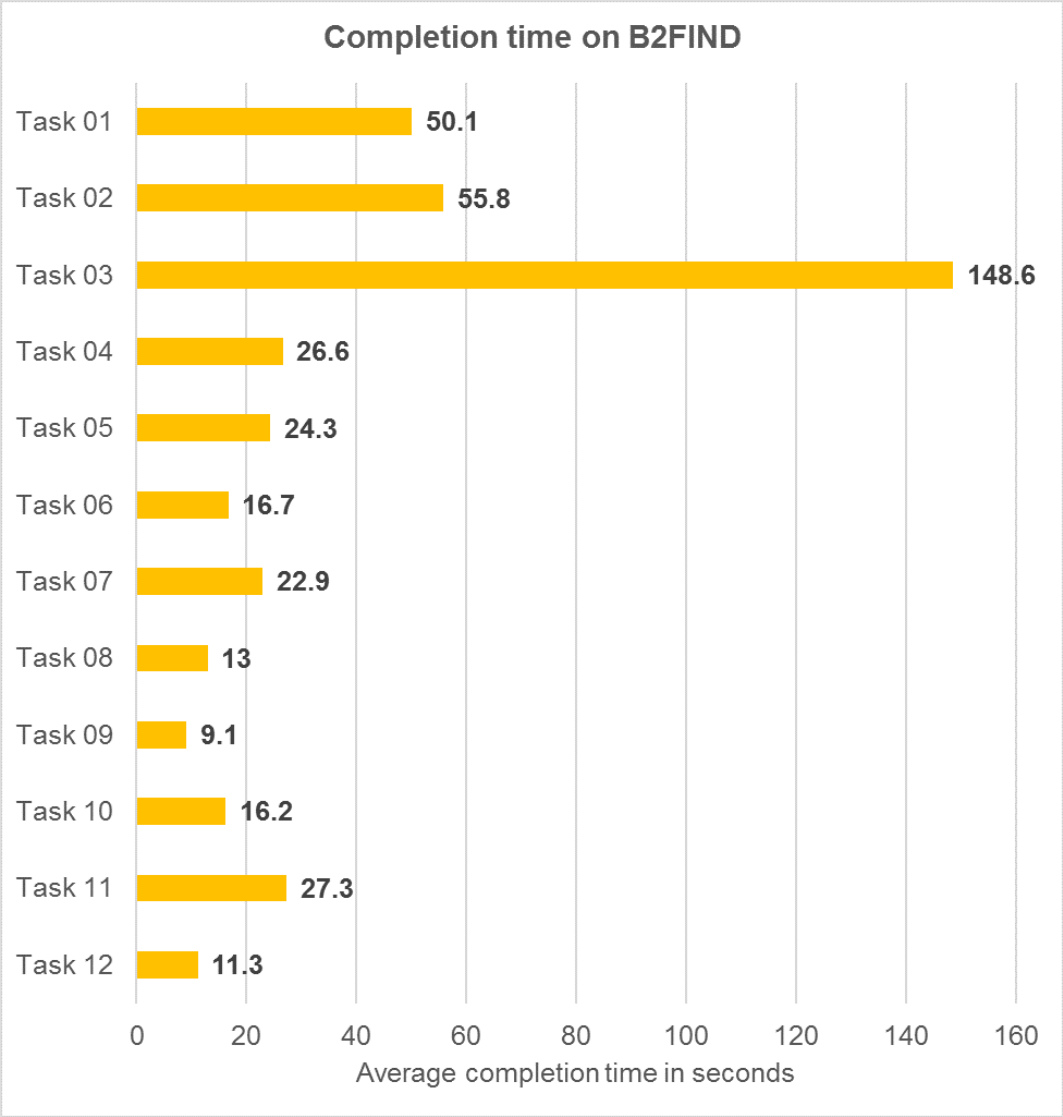

Figure 5.11: Completion time on B2FIND

apply button, random numbers appear inside start and end fields on the filter by time feature (see Figure 5.14). In fact, the random numbers shown in filter by time feature represents the time period, including date, hour, minute, and even seconds (see Figure 5.15). Finally, the user still needs to click the search button in order to apply the search filter (see Figure 5.14).

To solve this issue, a suggestion to use a calendar-like feature for filter by time is proposed since it can be understood more easily. Eight test users mention that usually the year field is enough when searching for research

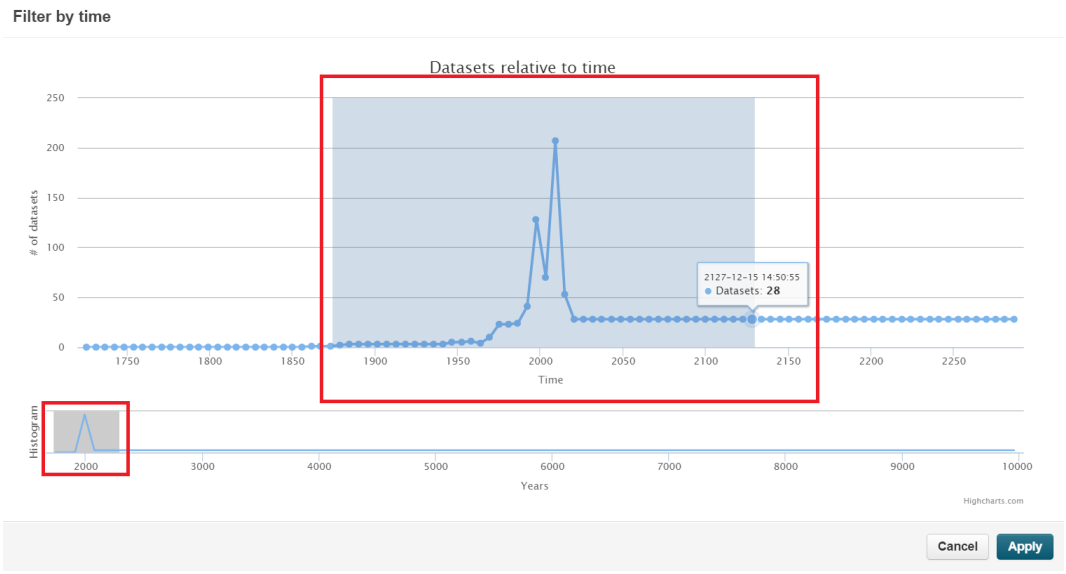

Figure 5.12: Filter by time on B2FIND

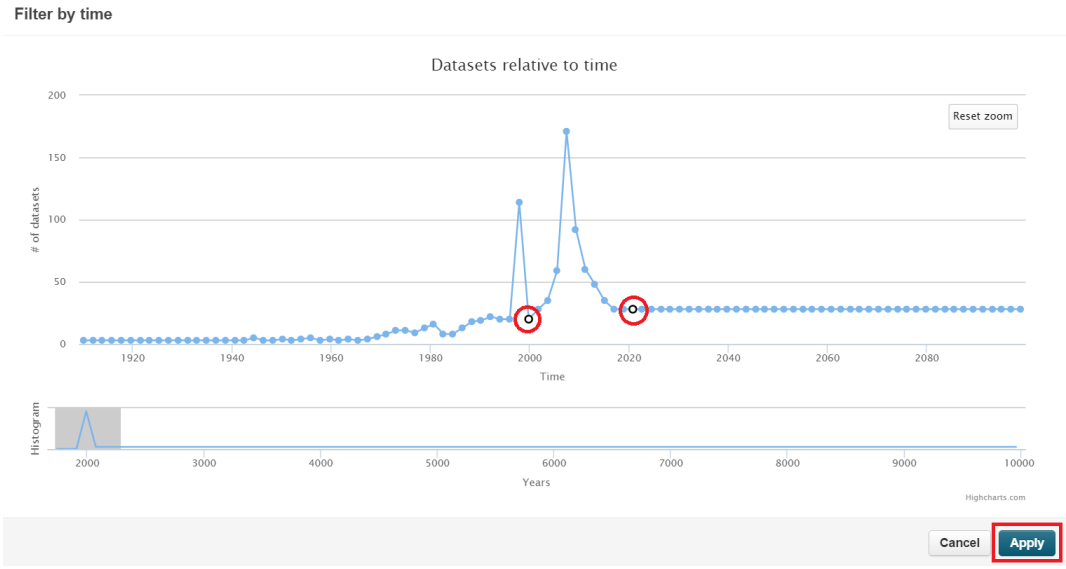

Figure 5.13: Filter by time on B2FIND

data. In some cases, the month field is also required. They also mention that the complete time format, including date, hour, minute, and second fields are not required. Therefore, a calendar-like feature is more suitable to be used for filter by time.

Fifteen percent of test users are not able to complete Task 07 (see Fig-

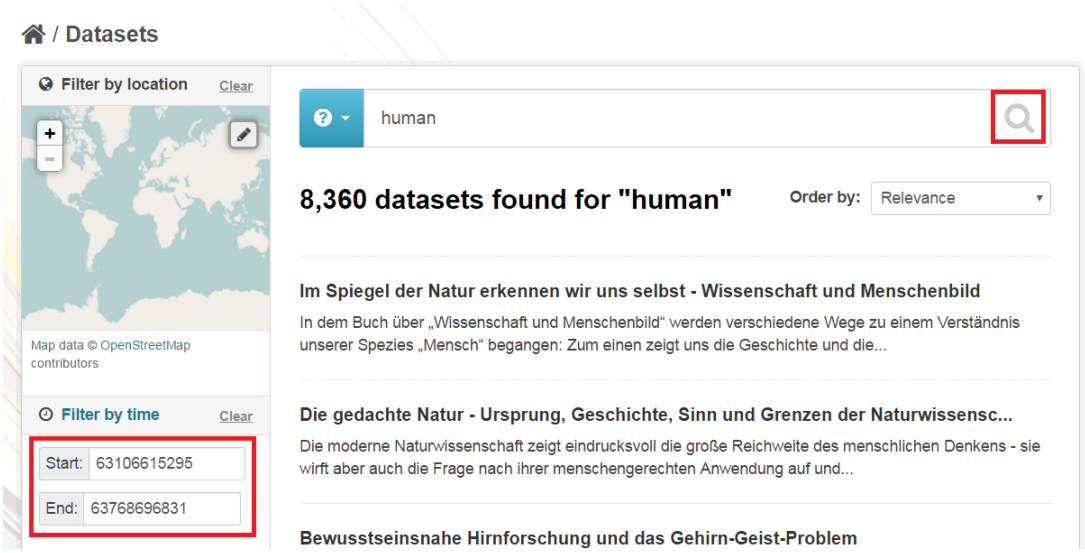

Figure 5.14: Filter by time on B2FIND

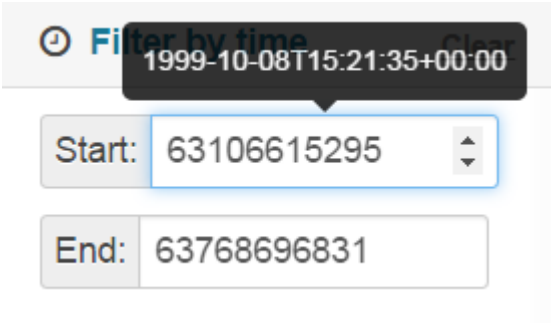

Figure 5.15: The meaning of random numbers in filter by time on B2FIND

ure 5.10). When test users have selected one of the available creators of datasets, the system shows internal server error. This issue is caused by problems with some creators of datasets. However, this problem does not occur to all creators of datasets. This issue can be avoided by simply clicking the back button on the web browser, then selecting other available creators. Nonetheless, the system should not only display internal server error when an error has occurred, but also include information about how to prevent or solve the error.

As what have been mentioned in Section 5.1.4, the inconsistency is one of usability issues in B2FIND. There is no consistency in how the search filters are being applied. Filter by location and time requires the user to click the apply button first before the search filter can be applied, while other search

filters are applied directly as soon as the user selected the available options. In addition to that, there is also no consistent way to select available options in search filters. Filter by tags, creator, and discipline allow the user to select multiple available options, while filter by community, language, and publisher allow the user to select one available option only. There should be a consistent way for applying search filters and selecting available options on search filters.

There are some new features that have been proposed for B2FIND. The new features have been proposed by test users during the usability testing and those features may increase the usefulness of B2FIND. Some of the proposed features are:

- Provide filter search result by types of data. This search filter is important since there are many types of research data and some researchers may need specific type of research data only.
- Display how many times the datasets have been cited. Number of citation can be used to show the credibility of the dataset and it would be useful for researchers to select appropriate datasets.
- Provide a button to clear all search filters. Currently, search filters that have been applied can be cleared one by one only. A button to clear all search filters that have been applied would be useful.

5.3 Results for B2DROP

There are seventeen tasks that need to be completed in B2DROP's task scenario (see Appendix C.2). Task 07 and Task 09 are the only tasks where not all of the test users were able to complete the tasks without help (see Figure 5.16).

Twenty-three percent of test users are not able to complete Task 07 without any help. There are two things that are contributing to this issue. The first thing is that the test users were not aware of drag-and-drop features on B2DROP. There is no notification about the possibility of doing drag-and-drop on the system. The second thing is that the process of moving files between folders is not straight-forward in B2DROP. In order to move files between folders, the user needs to bring the files to the upper directory first. After that, the user can drag-and-drop the files to the destination folder. Figure 5.17 also shows that it requires considerable amount of time for test users to complete Task 07.

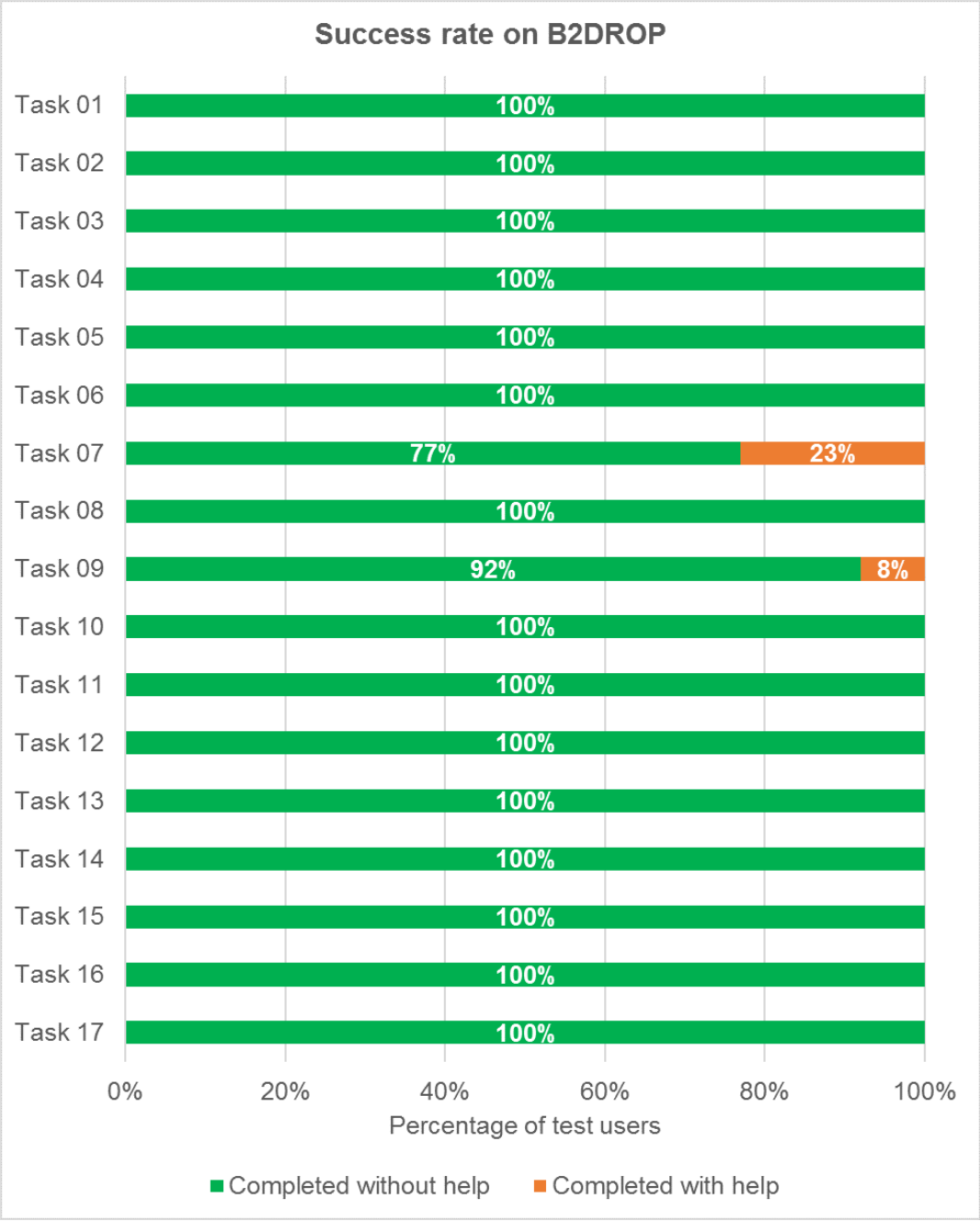

Figure 5.16: Success rate on B2DROP

There are eight percent of test users who are unable to complete Task 09 without help (see Figure 5.16). This issue is occurred since the features on B2DROP are well-hidden. Therefore, it is difficult to find the sharing

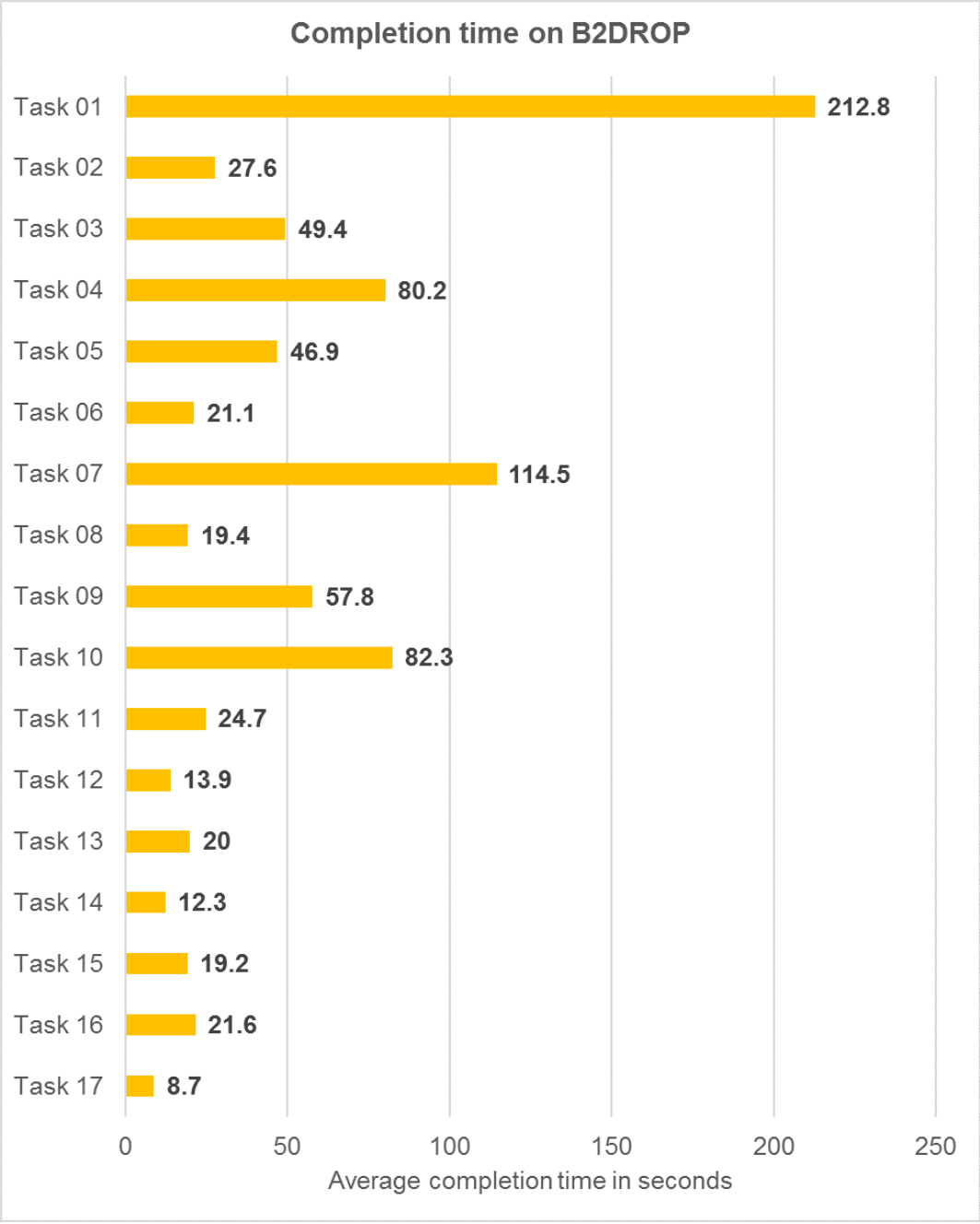

Figure 5.17: Completion time on B2DROP

feature on B2DROP (see Figure 5.18). In addition to that, two test users also mentioned that they do not understand the meaning of all icons on the user interface. In case of sharing features, the text will only appear after the

sharing has been applied (see Figure 5.19). To solve this issue, two test users suggest to put short text next to each feature on B2DROP. Therefore, the user would be able to find the features and understand the meaning of each icon on B2DROP more easily.

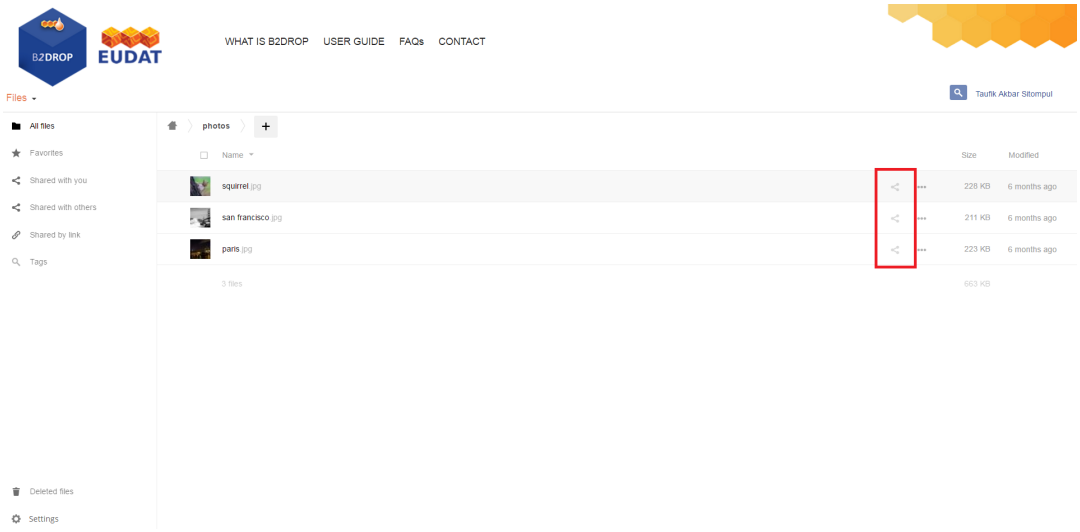

Figure 5.18: Before the sharing has been applied

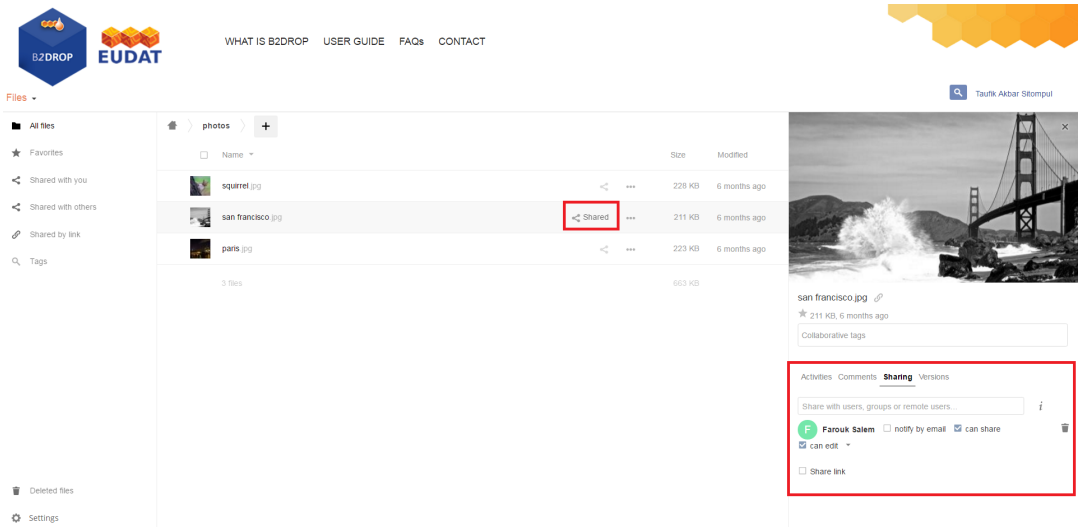

Figure 5.19: After the sharing has been applied

5.4 Results for B2SHARE

There are twelve tasks that need to be completed in B2SHARE's task scenario (see Appendix C.3). However, the first two tasks need to be completed in B2ACCESS. Therefore, there are only ten tasks in total that are actually involved with B2SHARE.

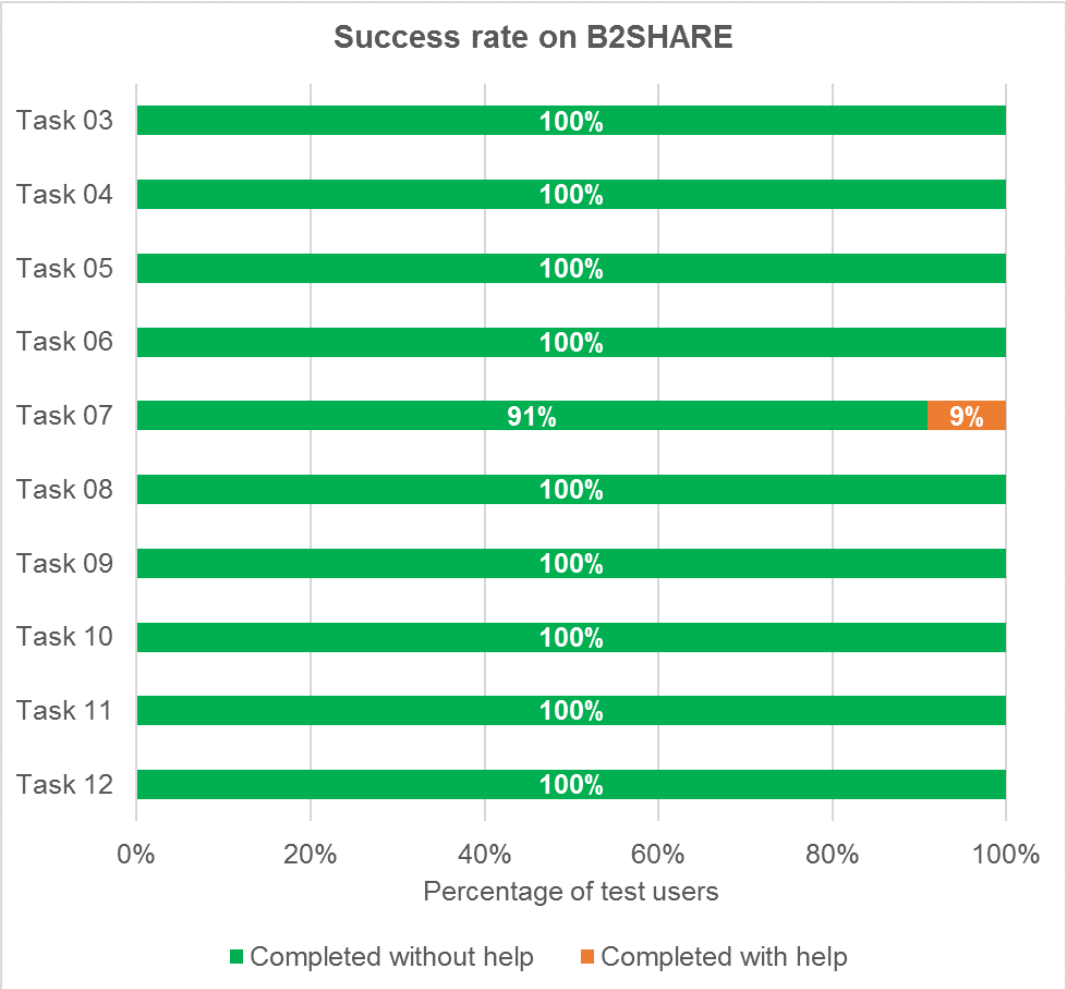

Figure 5.20: Success rate on B2SHARE

Task 07 is the only task which not all test users are able to complete without help (see Figure 5.20). The instruction in the Task 07 is "Find your uploaded dataset". There is a case where a test user is unable to find his uploaded dataset. The problem is occurred since it takes time for B2SHARE to index the uploaded dataset in order to make it searchable through the

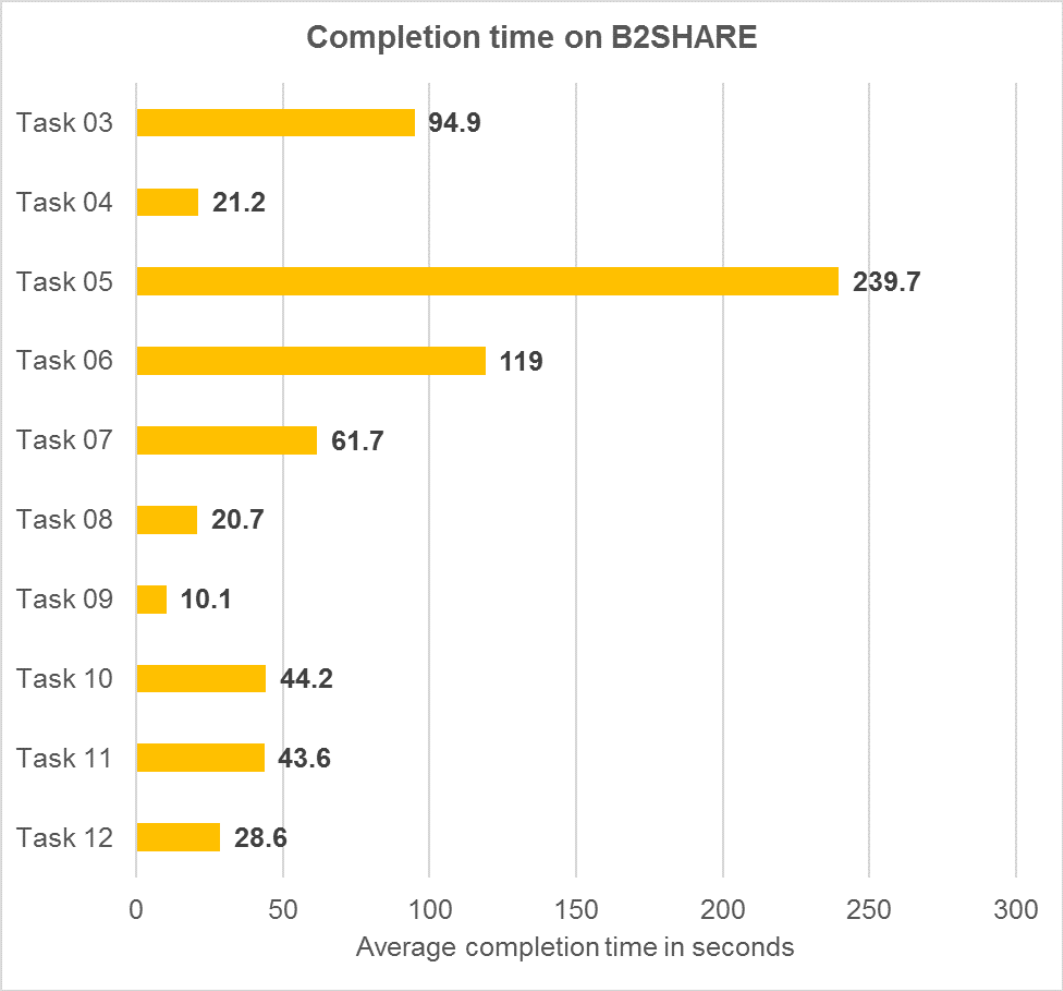

Figure 5.21: Completion time on B2SHARE

search form. However, the required time to index the uploaded dataset varies from one test session to another. During the usability testing, there are three test users who commented that the dataset should be available and searchable as soon as it has been uploaded. Figure 5.21 shows that it also requires considerable amount of time for the test users to find their uploaded datasets.

Although all of the test users are able to complete the rest of the test tasks without help, there are still many things that need to be improved on B2SHARE. Currently, there are some features on B2SHARE that are not functioning properly and the test users felt that the system is still far from complete. Following are some of notable features that are not functioning properly in B2SHARE:

- Sorting of search results is not functioning at all. The user interface of B2SHARE shows that the user can sort the search results based on several factors, such as alphabetically, most recent, most cited, and most relevant. However, none of those sorting features are functioning.
- Login fields still exist after the user has successfully logged in to B2SHARE. This problem makes the test users feel confused since the system is still showing the login fields, after successful login procedure (see Figure 5.22). The login fields should not appear anymore once the user has successfully logged in to B2SHARE.

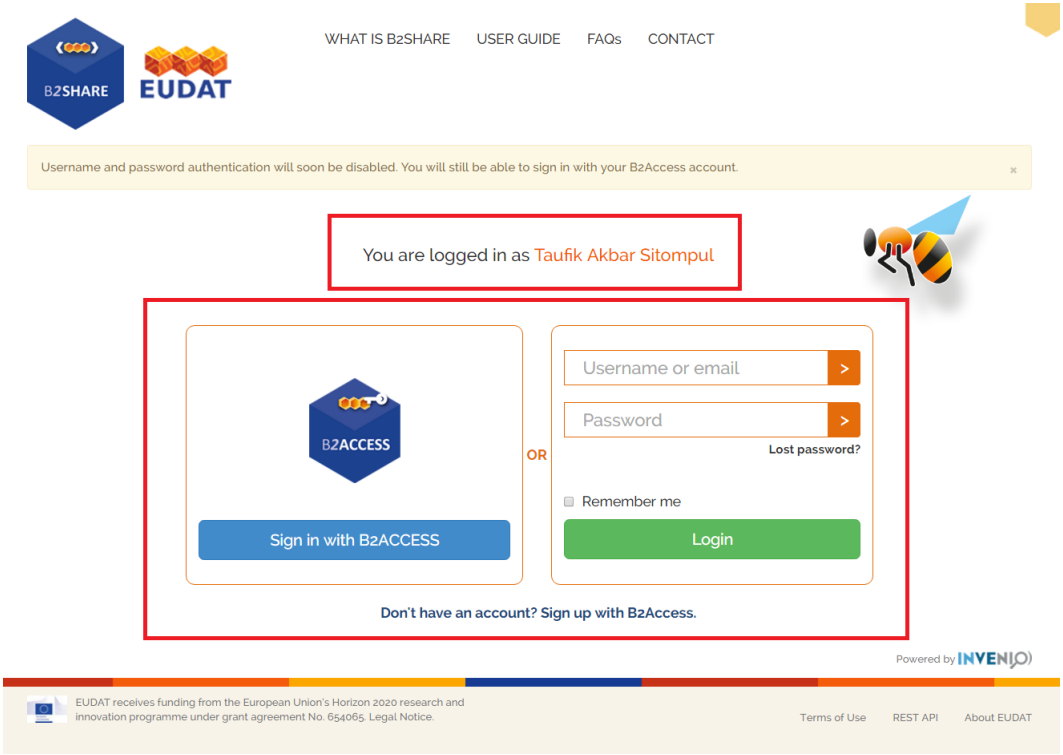

Figure 5.22: The user interface of B2SHARE when the user has successfully logged in.

- Finding similar records always displays an empty result. There is a feature in B2SHARE that can be used to find any dataset that has similar records to datasets that have been selected previously. However, this feature is not functioning since the system always displays empty results after searching the similar records.

5.5 Results for B2ACCESS

B2ACCESS is the only EUDAT service that does not have its own task scenario. However, the first two tasks in B2SHARE's task scenario need to be done on B2ACCESS (see Appendix C.3). Although only two tasks are related to B2ACCESS, those two tasks covers the main functionalities of B2ACCESS.

The results from the usability testing shows that the Task 01 is too difficult since all test users are not able to complete the task without help from the examiner (see Figure 5.23). It is also important to note that the Task 01 contains long instructions on how to complete the task, but test users are still not able to complete the task without help. Moreover, test users also describe their experience when creating an account on B2ACCESS as a confusing phase. Figure 5.24 shows that it also requires considerable amount of time for test users to complete Task 01.

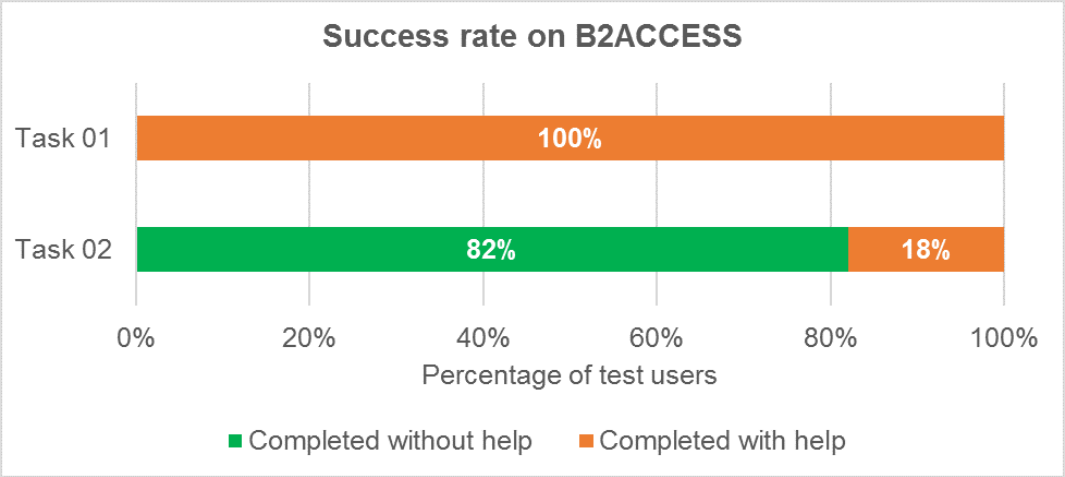

Figure 5.23: Success rate on B2ACCESS

Inability to complete Task 01 is caused by several usability problems, such as:

- B2ACCESS offers many options to create an account and displays all of them in one interface (see Figure 3.8). This makes the test users feel confused when choosing the option to create an account on B2ACCESS.
- There are too many steps that need to be taken into account to create an account on B2ACCESS. The registration process is not straightforward and it makes the test users do not really understand the whole registration process.

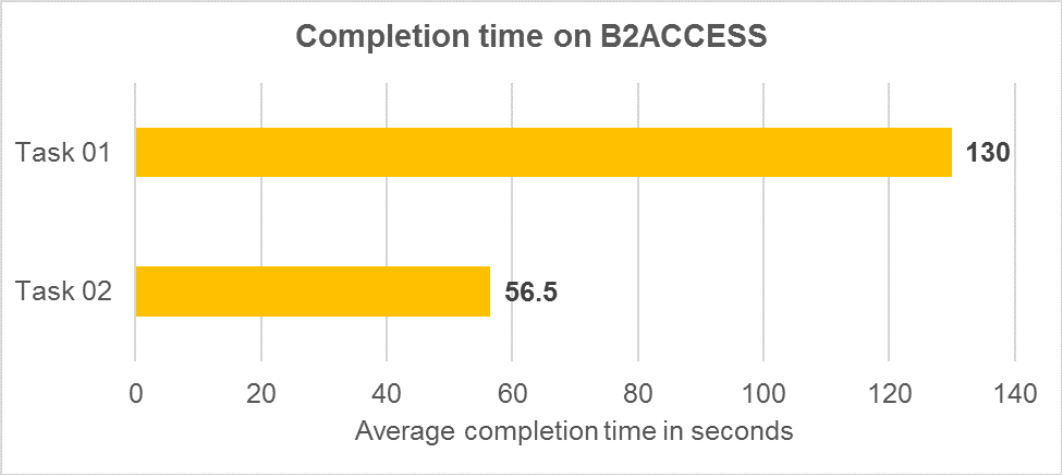

Figure 5.24: Completion time on B2ACCESS

- The dialogues that are shown on the system are not meaningful for the users (see Figure 5.25). The dialogues do not contain information that can be easily understood by the users.

Task 02 has similar, but much shorter process than Task 01. However, it does not make Task 02 easier at all for the test users. After completing Task 01, there are still eighteen percent of test users who are unable to complete Task 02 without any help (see Figure 5.23). Figure 5.24 shows that it also requires considerable amount of time for test users to complete Task 02. It should not take more than ten seconds for the test users to be able to log in to B2ACCESS, especially when they are already using their organizational credentials.

There are two main reasons that made two test users could not complete Task 02, such as:

- The long and complex registration process makes the test users forget what they have done on the system previously. Therefore, when they are trying to complete Task 02, they still have no idea on what should be done on the system.

- There are error messages as soon as the test users has clicked "Sign in with B2ACCESS" (see Figure 5.26). This makes test users feel confused since an error message appears when they have not done anything on the system. The error message also does not contain any information that can be used to help users to solve the error. When test users try to click "Try to continue", a new error message appears (see Figure 5.27).

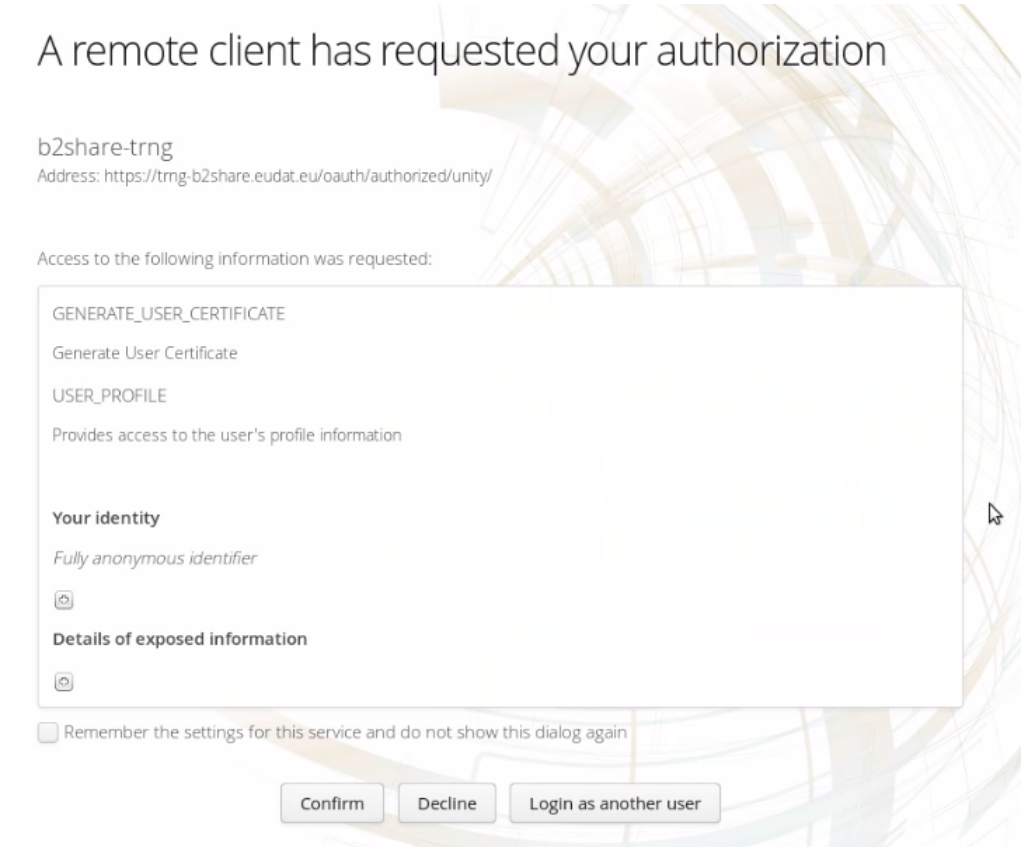

Figure 5.25: A system's dialogue on B2ACCESS

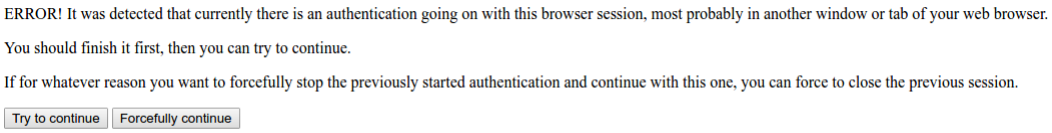

Figure 5.26: An error message on B2ACCESS

Based on these findings, there are three suggestions that have been proposed in order to improve the usability of B2ACCESS. Those three suggestions are:

- Provide meaningful messages on the system's dialogues. This is important to ensure that the users always understand things that are happening on the system.
- Shortening registration and login process. There are some steps in current registration and login processes that can be removed.

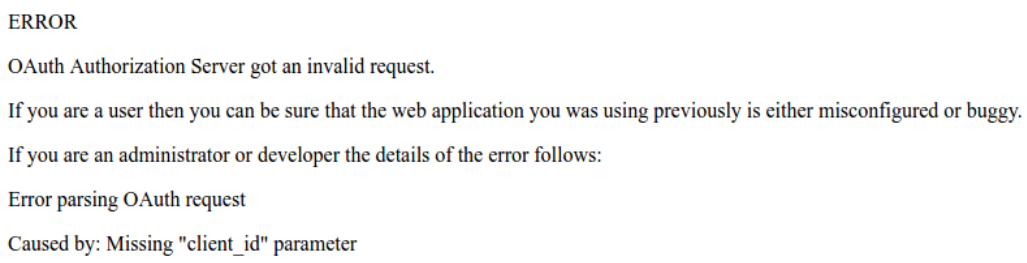

Figure 5.27: Another error message on B2ACCESS

- Include information about how to fix the error on the error messages. Good error messages do not only describe the error clearly, but also inform the users how to fix those errors.

5.6 Future work

Fixing the problems that have been found and implementing the suggestions that have been proposed from both UEMs do not guarantee that the future releases of EUDAT services will be free of usability issues. There might be new usability issues depending on how developers of EUDAT fixing the problems and implementing the suggestions. Therefore, it is highly recommended to conduct a similar evaluation for every major release of EUDAT services, especially if Aalto University decides to use EUDAT services as one of its research data management platforms.

If Aalto University decides to use EUDAT services, it is important to make sure that the researchers from all schools in Aalto University are involved in the future evaluations. Involving researchers from all schools is required to ensure that the results cover various requirements of research data management in Aalto University. The number of test users does not have to be large, but each school should have at least two representatives.

Involving developers of EUDAT in the future evaluations would be also beneficial for both the examiner and the developers. Involving developers of EUDAT allows the examiner to use different usability inspection methods, such as cognitive walkthrough, pluralistic walkthrough, formal usability inspection, and consistency inspection. Involving developers of EUDAT will also produce more suitable recommendations since no one knows EUDAT services better than the developers themselves. In addition to that, the developers of EUDAT do not have to wait until the final report has been

produced before start fixing the problems and implementing the suggestions. Fixing the problems and implementing the suggestions can be conducted as soon as the problems have been identified and the suggestions have been proposed, respectively.

Chapter 6

Conclusions

This thesis evaluates the state of usability and user experience of four EUDAT services: B2FIND, B2DROP, B2SHARE, and B2ACESS. From many available usability evaluation methods (UEMs), feature inspection and usability testing are used to evaluate EUDAT services. Feature inspection is preferred because of its straight-forwardness and short execution time. Usability testing is selected since involvement of test users is essential to discover usability problems.

Feature inspection is conducted by the author himself without involving any test user. For each feature on the EUDAT services, the author inspects its accessibility, understandability, and capability. There are thirteen researchers from various disciplines within Aalto University who participated in the usability testing. However, the goals of the usability testing have not been fully achieved since no researchers from School of Business were able to participate within the given schedule. Therefore, the proposed suggestions do not cover Aalto University's requirements completely.

Both quantitative and qualitative studies are used to analyze the results from both of the UEMs. Quantitative analysis, such as success rate and completion time are used to measure test users' performance when using EUDAT services. Quantitative analysis is also used to calculate test users' responses from the post-task questionnaires. On the other hand, qualitative analysis is used to describe what test users are thinking and feeling after using EUDAT services. In addition to that, qualitative analysis is also used to request modification on existing features and propose new features that would be useful to improve the usefulness of EUDAT services.

In total, both of the UEMs discover twenty-eight problems and propose ninety-three suggestions. Although the number of problems and suggestions are quite large, there are only a few overlapping results between both UEMs. Therefore, it can be concluded that feature inspection and usability testing

are complementing each other.

The results show that EUDAT services are quite applicable for Aalto University's research data management platform. Despite of existing usability problems in B2FIND and B2DROP, test users consider them as well enough. Test users state that B2FIND and B2DROP can fulfill the primary requirements of a search engine and a cloud storage service for research data, respectively. Although the success rate on B2SHARE is satisfying, test users state that B2SHARE feels buggy since some core functions, such as the advanced search and sorting features are not functioning properly, which makes B2SHARE look unreliable. B2ACCESS also requires many improvements. As one of the major improvement, it is essential to make registration and login process less complex and confusing, especially if the users are already using their institutional credentials on B2ACCESS.

It is highly recommended to conduct a similar evaluation for every major release of EUDAT services, especially if Aalto University decides to use EUDAT services as one of its research data management platforms. There might be new usability issues in newer versions depending on how developers of EUDAT are fixing the problems and implementing the suggestions. Involving developers of EUDAT would be beneficial as well since it allows the examiner to use different UEMs and the recommendations will be more suitable since no one knows EUDAT services better than the developers themselves. In addition to that, fixing problems and implementing suggestions can be conducted much earlier since the developers of EUDAT do not have to wait until the final report has been produced. Finally, it is also important to ensure that each school in Aalto University has at least two representatives. Therefore, the results of the evaluation will cover various requirements of research data management in Aalto University.

Bibliography

[1] William Albert and Thomas Tullis. *Measuring the user experience: Collecting, analyzing, and presenting usability metrics*. Interactive Technologies. Newnes, 2013. ISBN 9780124157927.

[2] Chris Andrzejczak and Dahai Liu. The effect of testing location on usability testing performance, participant stress levels, and subjective testing experience. *Journal of Systems and Software*, 83(7):1258–1266, 2010. URL http://dx.doi.org/10.1016/j.jss.2010.01.052.

[3] Nigel Bevan. Extending quality in use to provide a framework for usability measurement. In *Human Centered Design, First International Conference, HCD 2009, Held as Part of HCI International 2009, San Diego, CA, USA, July 19-24, 2009, Proceedings*, pages 13–22, 2009. URL http://dx.doi.org/10.1007/978-3-642-02806-9_2.

[4] Nigel Bevan. What is the difference between the purpose of usability and user experience evaluation methods. In *Proceedings of the Workshop UXEM*, volume 9, pages 1–4, 2009. URL http://nigelbevan.com/papers/What_is_the_difference_between_usability_and_user_experience_evaluation_methods.pdf.

[5] Randolph G. Bias. The pluralistic usability walkthrough: Coordinated empathies. In Jakob Nielsen and Robert L. Mack, editors, *Usability Inspection Methods*, chapter 3, pages 63–76. John Wiley & Sons, Inc., New York, NY, USA, 1994. ISBN 0-471-01877-5.

[6] John Brooke et al. Sus - A quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996. ISBN 9780748404605.

[7] Yuliana Puerta Cruz, César A. Collazos, and Toni Granollers. The thin red line between usability and user experiences. In *Proceedings of the XVI International Conference on Human Computer Interaction, Interacción 2015, Vilanova i la Geltrú, Spain, September 7-9, 2015*, pages 46:1–46:2, 2015. URL http://doi.acm.org/10.1145/2829875.2829915.

[8] Heather Desurvire, Debbie Lawrence, and Michael Atwood. Empiricism versus judgement: Comparing user interface evaluation methods on a new telephone-based interface. *SIGCHI Bulletin*, 23(4):58–59, 1991. URL `http://doi.acm.org/10.1145/126729.1056062`.

[9] Heather W. Desurvire. Faster, cheaper!! Are usability inspection methods as effective as empirical testing? In Jakob Nielsen and Robert L. Mack, editors, *Usability Inspection Methods*, chapter 7, pages 173–202. John Wiley & Sons, Inc., New York, NY, USA, 1994. ISBN 0-471-01877-5.

[10] Heather W. Desurvire, Jim M. Kondziela, and Michael E. Atwood. What is gained and lost when using evaluation methods other than empirical testing. In *Proceedings of the Conference on People and Computers VII*, HCI'92, pages 89–102, New York, NY, USA, 1993. Cambridge University Press. URL `http://dl.acm.org/citation.cfm?id=164592.164609`.

[11] ISO DIS. 9241-210: 2010. Ergonomics of human system interaction-part 210: Human-centred design for interactive systems. Technical report, 2010. URL `http://www.iso.org/iso/catalogue_detail.htm?csnumber=52075`.

[12] Sarah A Douglas. How to do usability testing. *Based on Douglas, Sarah (1995)." Conversation Analysis and Human-Computer Interaction Design". In PJ Thomas (Ed.), Social and Interactional Dimensions of Human-Computer Interfaces*, pages 184–203, 1998. URL `https://pdfs.semanticscholar.org/fa5e/ad49f3cbb7ea7dda6e62b14cea75a98ec420.pdf`.

[13] Joseph S Dumas and Janice Redish. *A practical guide to usability testing*. Intellect Books, 1999. ISBN 9781841500201.

[14] Kate Ehrlich and Janice Rohn. Cost justification of usability engineering: A vendor's perspective. In Randolph G. Bias and Deborah J. Mayhew, editors, *Cost-justifying Usability*, chapter 4, pages 73–110. Academic Press, Inc., Orlando, FL, USA, 1994. ISBN 0-471-01877-5.

[15] Laura Faulkner. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers*, 35(3):379–383, 2003. URL `http://dx.doi.org/10.3758/BF03195514`.

[16] Adrian Fernandez, Emilio Insfrán, and Silvia Abrahão. Usability evaluation methods for the Web: A systematic mapping study. *Information*

& Software Technology, 53(8):789–817, 2011. URL http://dx.doi.org/10.1016/j.infsof.2011.02.007.

[17] Jim Gray. Data management: Past, present, and future. *CoRR*, abs/cs/0701156, 2007. URL http://arxiv.org/abs/cs/0701156.

[18] Wayne D. Gray and Marilyn C. Salzman. Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-Computer Interaction*, 13(3):203–261, 1998. URL http://dx.doi.org/10.1207/s15327051hci1303_2.

[19] Cathy Gunn. An example of formal usability inspections in practice at Hewlett-Packard company. In *Human Factors in Computing Systems, CHI '95 Conference Companion: Mosaic of Creativity, Denver, Colorado, USA, May 7-11, 1995.*, pages 103–104, 1995. URL http://doi.acm.org/10.1145/223355.223451.

[20] Marion Hansen. Ten steps to usability testing. In *Proceedings of the 9th Annual International Conference on Systems Documentation, SIGDOC 1991, Chicago, Illinois, USA, 1991*, pages 135–139, 1991. URL http://doi.acm.org/10.1145/122778.122798.

[21] H. Rex Hartson, José C. Castillo, John T. Kelso, and Wayne C. Neale. Remote evaluation: The network as an extension of the usability laboratory. In *Conference on Human Factors in Computing Systems: Common Ground, CHI '96, Vancouver, BC, Canada, April 13-18, 1996, Proceedings.*, pages 228–235, 1996. URL http://doi.acm.org/10.1145/238386.238511.

[22] Marc Hassenzahl and Noam Tractinsky. User experience - A research agenda. *Behaviour & IT*, 25(2):91–97, 2006. URL http://dx.doi.org/10.1080/01449290500330331.

[23] Patrick A Holleran. A methodological note on pitfalls in usability testing. *Behaviour & Information Technology*, 10(5):345–357, 1991. URL http://dx.doi.org/10.1080/01449299108924295.

[24] Tasha Hollingsed and David G. Novick. Usability inspection methods after 15 years of research and practice. In *Proceedings of the 25th Annual International Conference on Design of Communication, SIGDOC 2007, El Paso, Texas, USA, October 22-24, 2007*, pages 249–255, 2007. URL http://doi.acm.org/10.1145/1297144.1297200.

[25] Andreas Holzinger. Usability engineering methods for software developers. *Commun. ACM*, 48(1):71–74, 2005. URL `http://doi.acm.org/10.1145/1039539.1039541`.

[26] James Hom. The usability methods toolbox handbook. 1998. URL `https://pdfs.semanticscholar.org/7f9b/8cdadcaae15736fc23a7e44cef843c9a4030.pdf`.

[27] Robin Jeffries, James R. Miller, Cathleen Wharton, and Kathy Uyeda. User interface evaluation in the real world: A comparison of four techniques. In *Conference on Human Factors in Computing Systems, CHI 1991, New Orleans, LA, USA, April 27 - May 2, 1991, Proceedings*, pages 119–124, 1991. URL `http://doi.acm.org/10.1145/108844.108862`.

[28] Patrick W Jordan. *Designing pleasurable products: An introduction to the new human factors*. CRC press, 2002. ISBN 978-0-7484-0844-3.

[29] Michael J. Kahn and Amanda Prail. Formal usability inspections. In Jakob Nielsen and Robert L. Mack, editors, *Usability Inspection Methods*, chapter 6, pages 141–171. John Wiley & Sons, Inc., New York, NY, USA, 1994.

[30] Laurie Kantner and Stephanie Rosenbaum. Usability studies of WWW sites: Heuristic evaluation vs. laboratory testing. In *The 15th Annual International Conference of Computer Documentation: Crossroads in Communication, SIGDOC 1997, Salt Lake City, Utah, USA, October 19-22, 1997*, pages 153–160, 1997. URL `http://doi.acm.org/10.1145/263367.263388`.

[31] Claire-Marie Karat, Robert Campbell, and Tarra Fiegel. Comparison of empirical testing and walkthrough methods in user interface evaluation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, pages 397–404, New York, NY, USA, 1992. ACM. URL `http://doi.acm.org/10.1145/142750.142873`.

[32] John Karat. User-centered software evaluation methodologies. *Handbook of human-computer interaction*, 2:689–704, 1997. ISBN 9780080532882.

[33] Khalid Majrashi, Margaret Hamilton, and A Uitdenbogerd. Multiple user interfaces and cross-platform user experience: Theoretical foundations. *SEA*, pages 43–57, 2015. URL `http://www.airccj.org/CSCP/vol5/csit53305.pdf`.

[34] Deborah J. Mayhew and Marilyn Mantei. A basic framework for cost-justifying usability engineering. In Randolph G. Bias and Deborah J. Mayhew, editors, *Cost-justifying Usability*, chapter 2, pages 9–43. Academic Press, Inc., Orlando, FL, USA, 1994. ISBN 0-12-095810-4.

[35] Rolf Molich and Jakob Nielsen. Improving a human-computer dialogue. *Commun. ACM*, 33(3):338–348, 1990. URL `http://doi.acm.org/10.1145/77481.77486`.

[36] Jakob Nielsen. Finding usability problems through heuristic evaluation. In *Conference on Human Factors in Computing Systems, CHI 1992, Monterey, CA, USA, May 3-7, 1992, Proceedings*, pages 373–380, 1992. URL `http://doi.acm.org/10.1145/142750.142834`.

[37] Jakob Nielsen. *Usability engineering*. Academic Press, 1993. ISBN 978-0-12-518405-2.

[38] Jakob Nielsen. Usability inspection methods. In *Human Factors in Computing Systems, CHI '95 Conference Companion: Mosaic of Creativity, Denver, Colorado, USA, May 7-11, 1995.*, pages 377–378, 1995. URL `http://doi.acm.org/10.1145/223355.223730`.

[39] Jakob Nielsen and Thomas K. Landauer. A mathematical model of the finding of usability problems. In *Human-Computer Interaction, INTERACT '93, IFIP TC13 International Conference on Human-Computer Interaction, 24-29 April 1993, Amsterdam, The Netherlands, jointly organised with ACM Conference on Human Aspects in Computing Systems CHI'93*, pages 206–213, 1993. URL `http://doi.acm.org/10.1145/169059.169166`.

[40] Jakob Nielsen and Rolf Molich. Heuristic evaluation of user interfaces. In *Conference on Human Factors in Computing Systems, CHI 1990, Seattle, WA, USA, April 1-5, 1990, Proceedings*, pages 249–256, 1990. URL `http://doi.acm.org/10.1145/97243.97281`.

[41] Miro Nurmela. Aalto data repository: Research data management, sharing and publishing in the world of data intensive science. Master's thesis, Aalto University, 2016. URL `https://aaltodoc.aalto.fi/handle/123456789/19414`.

[42] Helen Petrie and Christopher Power. What do users really care about?: A comparison of usability problems found by users and experts on highly

interactive websites. In *CHI Conference on Human Factors in Computing Systems, CHI '12, Austin, TX, USA - May 05 - 10, 2012*, pages 2107–2116, 2012. URL http://doi.acm.org/10.1145/2207676.2208363.

[43] Peter G. Polson, Clayton H. Lewis, John Rieman, and Cathleen Wharton. Cognitive walkthroughs: A method for theory-based evaluation of user interfaces. *International Journal of Man-Machine Studies*, 36(5):741–773, 1992. URL http://dx.doi.org/10.1016/0020-7373(92)90039-N.

[44] Johannes Reetz, Giovanni Morelli, Cristina Manzano, Kostas Kavoussanakis, Pascal Dugenie, David Vicente, Urpo Kaila, Claudio Cacciari, Heinrich Widmann, Florian Kaiser, Robert Verkerk, Dejan Vitlacil, Carl Johan Hakansson, and Pietari Hyväärinen. EUDAT D6.4 Final report on operations; v1.0. Technical report, May 2015. URL http://hdl.handle.net/11304/132bacae-f8a5-11e4-8a18-f31aa6f4d448.

[45] John Rieman, Marita Franzke, and David F. Redmiles. Usability evaluation with the cognitive walkthrough. In *Human Factors in Computing Systems, CHI '95 Conference Companion: Mosaic of Creativity, Denver, Colorado, USA, May 7-11, 1995.*, pages 387–388, 1995. URL http://doi.acm.org/10.1145/223355.223735.

[46] Sirpa Riihiaho. Experiences with usability evaluation methods. *Licentiate thesis. Helsinki University of Technology. Laboratory of Information Processing Science*, 2000. URL https://users.aalto.fi/~riihiaho/Riihiaho_thesis.pdf.

[47] Virpi Roto, Marianna Obrist, and Kaisa Väänänen-Vainio-Mattila. User experience evaluation methods in academic and industrial contexts. In *Proceedings of the Workshop UXEM*, volume 9. Citeseer, 2009. URL http://www.cs.tut.fi/~kaisavvm/UXEM09-Interact_ObristRotoVVM.pdf.

[48] Dominique L Scapin and Tomas Berns. Guest editorial usability evaluation methods. 1997. URL http://dx.doi.org/10.1080/014492997119770.

[49] Andrew Sears. Heuristic walkthroughs: Finding the problems without the noise. *Int. J. Hum. Comput. Interaction*, 9(3):213–234, 1997. URL http://dx.doi.org/10.1207/s15327590ijhc0903_2.

[50] Debora Shaw. Handbook of usability testing: How to plan, design, and conduct effective tests, by Jeffrey Rubin. *JASIS*, 47(3):258–259, 1996. URL `http://doi.acm.org/10.1145/169059.169166`.

[51] MW van Someren, Yvonne F Barnard, Jacobijn AC Sandberg, et al. *The think aloud method: A practical approach to modelling cognitive processes*. Academic Press, 1994. ISBN 0-12-714270-3.

[52] Tuula Teeri. Aalto University research data management policy. 2016. URL `http://goo.gl/ymNm16`.

[53] Katherine E. Thompson, Evelyn P. Rozanski, and Anne R. Haake. Here, there, anywhere: Remote usability testing that works. In *Proceedings of the 5th Conference on Information Technology Education, SIGITE 2004, 2004, Salt Lake City, UT, USA, October 28-30, 2004*, pages 132–137, 2004. URL `http://doi.acm.org/10.1145/1029533.1029567`.

[54] Henrik Thovtrup and Jakob Nielsen. Assessing the usability of a user interface standard. In *Conference on Human Factors in Computing Systems, CHI 1991, New Orleans, LA, USA, April 27 - May 2, 1991, Proceedings*, pages 335–341, 1991. URL `http://doi.acm.org/10.1145/108844.108937`.

[55] Mark van de Sanden, Christine Staiger, Claudio Cacciari, Roberto Mucci, Carl Johan Hakansson, Adil Hasan, Stephane Coutin, Hannes Thiemann, Benedikt von St. Vieth, and Jens Jensen. EUDAT D5.3 Final report on EUDAT services. Technical report, May 2015. URL `http://hdl.handle.net/11304/436879dc-f40a-11e4-ac7e-860aa0063d1f`.

[56] Sarah Waterson, James A. Landay, and Tara Matthews. In the lab and out in the wild: Remote web usability testing for mobile devices. In *Extended abstracts of the 2002 Conference on Human Factors in Computing Systems, CHI 2002, Minneapolis, Minnesota, USA, April 20-25, 2002*, pages 796–797, 2002. URL `http://doi.acm.org/10.1145/506443.506602`.

[57] Cathleen Wharton, John Rieman, Clayton Lewis, and Peter Polson. The cognitive walkthrough method: A practitioner's guide. In Jakob Nielsen and Robert L. Mack, editors, *Usability Inspection Methods*, chapter 5, pages 105–140. John Wiley & Sons, Inc., New York, NY, USA, 1994. ISBN 0-471-01877-5.

[58] Dennis Wixon, Sandra Jones, Linda Tse, and George Casaday. Inspections and design reviews: Framework, history and reflection. In Jakob

Nielsen and Robert L. Mack, editors, *Usability Inspection Methods*, chapter 4, pages 77–103. John Wiley & Sons, Inc., New York, NY, USA, 1994. ISBN 0-471-01877-5.

[59] Alan Woolrych and Gilbert Cockton. Why and when five test users aren't enough. In *Proceedings of IHM-HCI 2001 conference*, volume 2, pages 105–108. Citeseer, 2001. URL `https://fas-web.sunderland.ac.uk/~cs0gco/fiveusers.doc`.

Appendix A

Test Script

A.1 General introduction

Thank you very much for willing to participate as a test user of the usability testing. Before we get started, I would like to give you two short introductions. First of all, have you heard about Aalto University research data management policy?

(Skip this paragraph if the test user has heard about Aalto University research data management policy). Basically, Aalto University research data management policy is a policy to make research data management easier and the policy requires Aalto University to have platforms for its research data management. The policy also provides guidelines for researchers within Aalto University to publish their research data publicly.

The second short introduction is about EUDAT services. EUDAT is a funded project by EU and it offers common data services that cover the whole life cycle of research data, including storing, finding, managing, publishing, and preserving research data. Currently, there are six EUDAT services that can be used, but you will test four of them only. EUDAT is considered as one of the best platforms for research data management in Aalto University.

As you may already know, the purpose of this test is to test those four EUDAT services in order to see if EUDAT services are really suitable for researchers in Aalto University.

The test starts when I give you the list of task that needs to be completed in a specific EUDAT service. Please read the tasks carefully and feel free to ask me there is something that is unclear for you. Please do your best to complete the test tasks and do not be afraid to make mistakes. If you cannot complete the test tasks by yourself, you can refer to the manuals. If the manuals cannot help you, then I will help you to complete the specific test

task. While completing the test tasks, you are also asked to do think aloud, which means you need to say whatever comes to your mind. All kinds of comments are welcomed.

After completing the test tasks, you need to fill some questionnaires and there will be an open discussion afterwards where you can describe what you are thinking and feeling after using the specific EUDAT service. Your face, voice, and screen will be recorded during this test session. The video will not be seen by people outside this project. The video would be useful for me to do further analysis. Do you have any question so far?

A.2 Introduction and background questionnaire for each EUDAT service

(Read before the test user starts using a specific EUDAT service)

The first EUDAT service that you will use is B2FIND. B2FIND is a search engine for research data. Currently, there are more than 400.000 datasets from 15 communities that can be searched through B2FIND. Before we get started, there are two questions that I would like to ask you:

- Have you used any research data produced by other researchers?
- How did you obtain research data produced by other researchers?

The second EUDAT service that you will use is B2DROP. B2DROP is a cloud storage service that can be used to store and exchange data. B2DROP is very similar to Dropbox and Google Drive. Before we get started, there are two questions that I would like to ask you:

- How did you share research data within your research group?
- Have you used Dropbox, Google Drive, or other similar services?

The third and fourth EUDAT services that you will use are B2SHARE and B2ACCESS. B2SHARE is a platform to store and publish research data, while B2ACCESS is an authentication and authorization that can be used to log in to EUDAT services. However, B2ACCESS can be used to log in to B2SHARE only at the moment. You can use your Aalto University's credentials to log in to B2ACCESS. Before we get started, there are two questions that I would like to ask you:

- What kind of research data that has been produced in your research group?
- Have you published any research data?

Appendix B

Test users' profile

Test User	Sex	Position	School	OS/ Browser	Experience
#1	Male	Doctoral student	SCI	Linux/ Firefox	He uses only generated research data from his research group. He has used Dropbox for many years. He never shares research data publicly.
#2	Male	Post-doctoral researcher	SCI	Linux/ Firefox	Sometimes he uses research data from open data portal. He has used Dropbox and Google Drive for a long time. He never shares research data publicly.
#3	Male	Post-doctoral researcher	SCI	Windows/ Chrome	Sometimes he uses research data from specific collaborators. He has used Dropbox for a long time. He never shares research data publicly.
#4	Male	Doctoral student	ENG	Linux/ Firefox	Sometimes he uses research data from corporate partners. He has used Dropbox and Google Drive for a long time. He never shares research data publicly.

#5	Male	Doctoral student	ENG	Linux/ Firefox	He uses only generated research data from his research group. He has used Dropbox for a long time. He never shares research data publicly.
#6	Male	Post-doctoral researcher	SCI	Linux/ Opera	He uses only generated research data from his research group. He has used Dropbox and Google Drive for a long time. He never shares research data publicly.
#7	Male	Doctoral student	ARTS	Mac/ Chrome	He uses only his own research data. He has used Dropbox, Google Drive, and OwnCloud for a long time. He never shares research data publicly.
#8	Male	Post-doctoral researcher	SCI	Linux/ Firefox	He uses only generated research data from his research group. He has used Google Drive and OwnCloud for a long time. He never shares research data publicly.
#9	Male	Research assistant	SCI	Linux/ Chrome	Sometimes he uses research data from open data portal He has used Dropbox and Google Drive for a long time. He never shares research data publicly.
#10	Male	Post-doctoral researcher	HIIT	Linux/ Firefox	Sometimes he uses research data from specific collaborators. He has used Dropbox, Google Drive, and OwnCloud for a long time. He has shared research data publicly.

#11	Female	Professor	ARTS	Mac/ Safari	She uses only generated research data from her research group. She has used Dropbox for a long time. She has shared research data publicly.
#12	Male	Post-doctoral researcher	SCI	Linux/ Firefox	He uses only generated research data from her research group. He has used Dropbox and Google Drive for a long time. He never shares research data publicly.
#13	Male	Post-doctoral researcher	SCI	Linux/ Firefox	Sometimes he uses research data from open data portal. He has used Dropbox and Google Drive for a long time. He has shared research data publicly.

Abbreviations: SCI = School of Science; ENG = School of Engineering; ARTS = School of Arts, Design, and Architecture; HIIT = Helsinki Institute for Information Technology.

Appendix C

List of test tasks for the test users

C.1 List of test tasks to be completed on B2FIND

- Task 01 - Search a dataset by using common keywords that interest you. Make sure the search results are huge, for example more than 5000 datasets.
- Task 02 - Filter the search results by location and select the whole Europe.
- Task 03 - Filter the search result by time and select between 2000 and 2020.
- Task 04 - Filter the search result by publication year and select between 2010 and 2015.
- Task 05 - Filter the search result by selecting one of available communities. Skip this step if there are no available communities.
- Task 06 - Filter the search result by selecting one of available tags. Skip this step if there are no available tags.
- Task 07 - Filter the search result by selecting one of available creators. Skip this step if there are no available creators.
- Task 08 - Filter the search result by selecting one of available disciplines. Skip this step if there are no available disciplines.

- Task 09 - Filter the search result by selecting one of available languages. Skip this step if there are no available languages.
- Task 10 - Filter the search result by selecting one of available publishers. Skip this step if there are no available publishers.
- Task 11 - Select one of the dataset and check the metadata.
- Task 12 - Click and open the source of the dataset.

C.2 List of test tasks to be completed on B2DROP

- Task 01 - Create a new account on B2DROP.
- Task 02 - Create at least 2 folders.
- Task 03 - Rename folders that you have created.
- Task 04 - Upload at least 3 files into a folder that you have created.
- Task 05 - Upload a same file twice on the same folder and do versioning.
- Task 06 - Rename files that you have uploaded.
- Task 07 - Move files that you have uploaded to another folder.
- Task 08 - Set a file as favourite.
- Task 09 - Share a file with another B2DROP user. Share the file with Taufik Akbar Sitompul.
- Task 10 - Select another file, share it by link and set an expiry date. Share the file with taufik.sitompul@aalto.fi.
- Task 11 - Unshare a file.
- Task 12 - Move a folder into another folder.
- Task 13 - Delete files.
- Task 14 - Delete folders.
- Task 15 - Restore a file or a folder from trash bin.
- Task 16 - Delete files or folders from trash bin permanently.
- Task 17 - Sign out from B2DROP.

C.3 List of test tasks to be completed on B2SHARE

- Task 01 - Create a new account on B2SHARE. Click the register button and then select register with B2ACCESS. Search for Aalto University and login with your Aalto University's credentials.
- Task 02 - Sign in to B2SHARE and then select sign in with B2ACCESS.
- Task 03 - Click deposit and select a dataset to be uploaded.
- Task 04 - Select an appropriate domain for your dataset.
- Task 05 - Fill basic details of the dataset.
- Task 06 - Add more details to the dataset and deposit your dataset.
- Task 07 - Search and select a dataset that interests you.
- Task 08 - Download the dataset.
- Task 09 - Export selected dataset into a citation format, e.g. BibTex, MARCXML, EndNote, etc.
- Task 10 - Rate and give a comment to the selected dataset.
- Task 11 - Find your uploaded dataset.
- Task 12 - Sign out from B2SHARE.

Appendix D

Post-task questionnaires

D.1 Modified System Usability Scale (SUS)

No	Statement	Strongly Disagree	Disagree	Agree	Strongly Agree	Not Applicable
1	I think that I would like to use this system frequently					
2	I found that this system is simple					
3	I found that this system is easy to use					
4	I think I would need the support of a technical person to be able to use this system					
5	I found that various functions in this system are well-integrated					
6	I think this system is consistent					
7	I imagine that most people would learn to use this system very quickly					
8	I found that this system is very troublesome to use					
9	I felt very confident using this system					

10	I think I need to learn a lot of things before I could get going with this system					

D.2 Modified Computer System Usability Questionnaire (CSUQ)

No	Statement	Strongly Disagree	Disagree	Agree	Strongly Agree	Not Applicable
1	I can effectively complete my work using this system					
2	I am able to complete my work quickly using this system					
3	I am able to efficiently complete my work using this system					
4	I feel comfortable using this system					
5	It was easy to learn to use this system					
6	I believe I became productive quickly using this system					
7	This system gives error messages that clearly tell me how to fix problems					
8	Whenever I make a mistake using the system, I can recover easily and quickly					
9	Additional information (such as online help, on-screen messages, and other documentation) provided with this system is clear					

10	It is easy to find the information I needed					
11	The information provided for this system is easy to understand					
12	The organization of information on the screen is clear					
13	The interface of this system is pleasant					
14	I like using the interface of this system					
15	This system has all functions and capabilities that I expect it to have					

D.3 Modified Questionnaire for User Interaction Satisfaction (QUIS)

No	Overall Reaction to The Software	Strongly Disagree	Disagree	Agree	Strongly Agree	Not Applicable
1	This system is satisfying					
2	This system is stimulating					
3	This system is flexible					
No	Screen	Strongly Disagree	Disagree	Agree	Strongly Agree	Not Applicable
4	Reading characters on the screen is easy					
5	Sequence of screens is clear					
No	Terminology and System Information	Strongly Disagree	Disagree	Agree	Strongly Agree	Not Applicable
6	Use of terms throughout the system is consistent					

7	Terminology is always related to the task					
8	Position of messages on the screen is consistent					
9	Prompts for input are clear					
10	This system always inform about its progress					
No	Learning	Strongly Disagree	Disagree	Agree	Strongly Agree	Not Applicable
11	Exploring new features by trial and error is easy					
12	Remembering names and use of commands are easy					
13	Performing tasks is always straight-forward					
14	Help messages on the screen are helpful					
No	System Capabilities	Strongly Disagree	Disagree	Agree	Strongly Agree	Not Applicable
15	This system's speed is fast enough					
16	This system is reliable					
17	This system tends to be quiet					
18	This system is designed for all levels of users					

Appendix E

List of problems and suggestions

Table E.1: Usability evaluation methods (UEMs)

Code	UEM	Description
FI	Feature inspection	The first UEM thas been conducted. The author inspects each available feature on EUDAT services. There are no test users involved in this UEM.
UT	Usability testing	The second UEM that has been conducted. There are 13 researchers who have participated as test users.

Table E.2: Severity level of found problems

Code	Severity	Description
CA	Catastrophe	This problem should be fixed before the system can be released.
MA	Major problem	This problem should be given high priority to be fixed.
MI	Minor problem	This problem can be given low priority to be fixed.
CO	Cosmetic problem	This problem does not need to be fixed, unless there is extra time.

Table E.3: Priority level of proposed suggestions

Code	Priority	Description
UR	Urgent	This suggestion should be implemented immediately.
HI	High	This suggestion should be given high priority to be implemented.
MO	Moderate	This suggestion can be implemented later.
LO	Low	This suggestion does not have to be implemented, unless there is extra time.

Table E.4: Status of reported problems and suggestions

Code	Priority	Description
FIX	Fixed	This problem has been fixed by developers of EUDAT services.
IMP	Implemented	This suggestion has been implemented by developers of EUDAT services.
TIC	Taken into consideration	Developers of EUDAT services are considering to fix/implement this problem/suggestion.
REM	Removed	This problem will not be fixed since the related feature will be removed in the next version of EUDAT services.
REJ	Rejected	This suggestion has been rejected to be implemented by developers of EUDAT services.

E.1 B2FIND

E.1.1 List of problems found on B2FIND

No	Problem	Description	UEM	Severity	Status
1	Activity stream does not display anything.	When a dataset has been selected, the tab for activity stream does not display anything.	FI	MI	TIC

2	Related tab always displays no related item.	When a dataset has been selected, the related tab always displays "no related item". Although, according to the search results, the specific dataset has similarities with other datasets.	FI	MI	TIC
3	Internal server error is occurred when selecting one of available creators.	Internal server error is occurred when test users filtering the search results according to the creator. However, this problem is occurred with some specific creators only.	UT	MI	TIC
4	Sometimes, filter by time does not show any graph.	Sometimes, when a test user has clicked filter by time, there is no graph shown on the system.	UT	MA	TIC

E.1.2 List of proposed suggestions for B2FIND

No	Suggestion	Description	UEM	Priority	Status
1	Simplify the process to do filter by time or use calendar-like feature for filter by time.	Currently, the process to do filter by time is quite complex. There are uncommon operations, many steps that need to be taken, and the filter is not applied right away. Users should not need to see user guides first in order to do filter by time. Filter by time should be using year only instead of complete date. Calendar-like feature can make filtering by time easier.	FI, UT	CR	TIC

2	Provide drop-down list of continents or countries for filter by location.	It might be easier to do filter by location by using a drop-down list of continents or countries that can be selected.	UT	MO	TIC
3	Provide filter search result by types of data.	There are many types of data that can be searched on B2FIND. A feature for filtering search results according to the types of data, such as image, text, video, and audio would be useful.	UT	MO	TIC
4	Implement one consistent way to apply the search filters.	Some search filters are applied directly and some are not applied directly. Sometimes, test users were not sure whether the search filter has been applied or not.	UT	HI	TIC
5	Display how many times the datasets have been cited.	The number of how many times a dataset has been cited may reflect the credibility of that dataset. This would be useful for researchers to select appropriate datasets.	UT	MO	TIC
6	Implement a consistent way to select available options on search filters.	Some search filters allow the user to select more than one available options and some allow the user to select one of available options only. It is unclear whether a user can select more than one available options or not.	UT	MO	TIC
7	Possibility to select more than one languages on filter by language.	Some countries have more than one national languages and the research data may be written in various languages. It should be possible to select more than one languages on filter by language.	UT	LO	TIC

8	Implement a consistent way to display available languages on filter by language.	Currently, there are different ways to display available languages (for example en, EN, and English) on filter by language. There should be a consistent way to display available languages.	UT	LO	TIC
9	Provide filter search result by types of license.	Each dataset is using different license. A feature for filtering search results according to the types of license would be useful. Researchers can find datasets according the most suitable licenses for them.	UT	LO	TIC
10	Move the search help to the navigation menu.	Currently, the search help is located next to the search field. The search help should not be located next to the search field to avoid misunderstanding. For example, the search help can be placed on the navigation menu.	UT	LO	TIC
11	Provide a small icon to show the availability of the dataset.	Some datasets are publicly available and some datasets have restricted access. It would be useful if B2FIND can mark or highlight which dataset is publicly available and which dataset needs to be requested first.	UT	MO	TIC
12	Move filter by discipline to higher position.	It makes more sense to do filter the search results by disciplines first before filter by tags and creator. Filter by discipline should be placed before or after filter by communities.	UT	MO	TIC

13	Possibility to search datasets from existing open data portal.	To increase the number of datasets that can be searched on B2FIND, it should be possible to search research data from existing open data portals, e.g., EU Open Data Portal.	UT	LO	TIC
14	Provide a button to clear all search filters.	Currently, the users need to clear search filters that have been applied one by one. There should be a button to clear all search filters that have been applied.	UT	HI	TIC

E.2 B2DROP

E.2.1 List of problems found on B2DROP

No	Problem	Description	UEM	Severity	Status
1	Change the contact's link on the navigation menu.	On the navigation menu, the contact's link should go to Contact & Support's page instead of sending a direct e-mail to EUDAT's support.	FI	MI	FIX
2	Problem with signing in after signed out.	There is a problem to sign in if the user has successfully signed out previously. To solve this problem, the user needs to close and re-launch the browser.	FI	MA	FIX
3	Mention email instead of username in the login field.	Currently, it is written username on the login field. This is confusing since what is needed to log in to B2DROP is email and not username.	UT	MA	FIX

4	The password on share by link disappears automatically.	On share by link, the password disappears automatically after the users have filled the password. The password should not disappear once it has been filled.	UT	MA	TIC
5	The email on share by link disappears automatically.	On share by link, the email disappears automatically when users clicked somewhere else. The email should not disappear once it has been filled.	UT	MA	TIC
6	Terms of use of B2DROP should be clickable and readable.	When creating an account, users need to accept terms of use of B2DROP. However, the terms of use of B2DROP is not clickable and readable.	UT	MA	TIC

E.2.2 List of proposed suggestions for B2DROP

No	Suggestion	Description	UEM	Priority	Status
1	Provide user guides for B2DROP.	Although EUDAT B2DROP is very easy to use, user guides are always a useful thing to provide, especially for new users.	FI	HI	IMP
2	Provide a button to change user's full name and email.	Although the changes for full name and email are made automatically, there should be a button to confirm the changes of full name and email. It provides sense of control to the users.	FI	LO	REJ
3	Provide indexing or drop-down feature for FAQ.	The list of questions should have drop-down feature. The answer will appear when the user has clicked the question.	FI	LO	TIC

4	Provide video player to preview video-type files.	Video player is a nice feature to have since photo and PDF viewer are already available in the system. Therefore, user can check the video first before downloading it.	FI	MO	IMP
5	Provide button or icon to move folders and files.	Although folders and files can be moved by using drag and drop, it is also nice to provide button or icon for moving files or folders. It makes the users aware of the system capabilities. Moreover, some users prefer to move files and folders manually.	FI, UT	HI	TIC
6	Provide feature to modify and delete multiple files or folders at one time.	Although there is a toggle feature in the system, user is still unable to modify or remove multiple les at the same time. It is a nice feature to have since some users may have plenty of files or folders in their storage.	FI	MO	IMP
7	Display pop-up warning before deleting files permanently.	Since users tend to make mistakes, there should be a pop-up warning to ask for confirmation from the user before deleting the files permanently.	FI, UT	HI	TIC
8	Display the contributor on the shared file and shared folder.	In case of shared file and shared folder, the system should display who is the uploader or the last modifier.	FI	MO	IMP
9	Provide a feature to give comment on the uploaded file.	In case of shared file, the system should allow users to give their comments on the uploaded files. It can encourage user collaboration and make communication among users becomes easier.	FI	MO	IMP

10	Sign in to B2DROP via B2ACCESS.	Instead of creating a new account, the user should be able to sign in to B2DROP by using the same account that has been registered through B2ACCESS.	FI, UT	HI	TIC
11	Provide file directory and show indication when users are dragging and hovering files or folders to destination folder.	Moving files are not straightforward. Users cannot move files or folders to destination folder directly.	UT	UR	TIC
12	Provide a button to confirm the sharing on share with other users.	On share with other users, the sharing should not be done automatically after selecting the user to be shared with. There should be a confirmation to confirm the sharing and then the sharing is applied.	UT	UR	TIC
13	Expose details, rename, download, and rename features.	Currently, details, rename, download, and rename features are hidden and they are difficult to be found since the system is using pull-down menu to display them. Those features can be displayed directly since there is also enough empty space on the interface.	UT	HI	TIC
14	Possibility to upload folders to B2DROP.	Currently, users can upload only files to B2DROP. However, some researchers have structured folders and files. It would be useful if users can upload the whole folder to B2DROP	UT	HI	TIC

15	The registration and login pages should have the same design as other pages.	Currently, the registration and login pages have different design compared to other pages. This made test users were thinking that they have been redirected into wrong pages.	UT	HI	TIC
16	File should not be automatically downloaded or previewed when the filename is clicked.	Currently, files stored in B2DROP will be automatically downloaded or previewed if the filename is clicked by the users. The file should not be downloaded or previewed automatically when the filename is clicked.	UT	UR	TIC
17	Shorten the verification code or use verification link only.	A verification email is sent automatically after users have registered an account on B2DROP. The email contains verification code and verification link. However, the verification code is very long and the verification link has little visibility since it is located on the bottom of the email. Most test users were not aware of the existence of the verification link.	UT	MO	TIC
18	Provide a confirmation to unshare file or folder.	Currently, the unshare will be applied directly when users unchecked the share by link or clicked the unshare button. There should be a confirmation dialogue to confirm the unshare, before the unshare is being applied.	UT	MO	TIC

19	Provide short text next to the icons.	Some test users did not know the meaning of some icons on B2DROP. Providing short text next to each icon would be useful to solve this issue.	UT	MO	TIC
20	Provide personalized link for sharing by link.	Sometimes, users want to share a file with different people. By using personalized link, users can decide which link that they want to keep and which link that they want to remove.	UT	LO	TIC
21	Provide link or icon to download B2DROP's client on the main page.	Currently, users need to go to personal page first if they want to download and install B2DROP's client. Providing a link or an icon on the main page that can be used to download B2DROP's client would be useful.	UT	LO	TIC
22	Move the log out button to the right top of the interface.	Currently, the logout button will appear when users clicked their username first. It would be better if the logout button always available on the right top of the interface.	UT	LO	TIC
23	Provide information regarding how long old versions of uploaded data will be stored in B2DROP.	It is not clear how long B2DROP will store old versions of uploaded data. There should be information to address this issue. For example, Dropbox stores old versions of uploaded data for 30 days.	UT	MO	TIC
24	Provide clear distinction between share with users and share by link.	It is not clear which one is sharing with other users and which one is sharing by link. B2DROP should give more emphasize the differences between share with other users and share by link.	UT	MO	TIC

25	Provide information regarding the limitation of B2DROP.	A test user failed to upload files larger than 2 GB. There should be information regarding the limitation of B2DROP.	UT	UR	TIC
26	List the current and the old versions of uploaded data.	Currently, B2DROP lists only the old versions of uploaded data. To avoid confusion, the list should also include the current and the old versions of uploaded data.	UT	LO	TIC
27	Possibility to delete old versions of uploaded data.	B2DROP should allow users to decide which version of uploaded data to be deleted since it consumes storage.	UT	LO	TIC
28	Possibility to rename files from the right panel.	Currently, a right panel appears when users clicked the file. In addition to current feature, it would be useful if users can rename the file from the right panel.	UT	LO	TIC
29	Provide clear distinction between can edit and can change on share with users.	When users have shared a file with another user, there are two options that are quite similar: can edit and can change. However, there is no clear difference between those two features.	UT	MO	TIC
30	Provide information regarding how secure and safe B2DROP is.	Security and privacy are one of the main considerations for researchers. B2DROP should provide some information regarding security and privacy of its service.	UT	UR	TIC
31	Provide different icon for README file.	Since B2DROP allows users to make text files, it would be useful if B2DROP has different icon for readme file in order to give more emphasize on it.	UT	LO	TIC

32	Do not use automatic translation on B2DROP.	There was a test user who used the chinese translation of B2DROP, but the translation was difficult to understand, even for a Chinese. B2DROP should not use automatic translation.	UT	MO	TIC
33	Possibility to edit document files online.	It would be useful if users can edit some document files online, for example, .doc, .xls, and .ppt files.	UT	MO	TIC
34	Displaying details, rename, download, and delete features by doing right click.	In addition to the existing mechanism, it would be useful if some features like details, rename, download, and delete can be shown by doing right click.	UT	MO	TIC
35	Provide information regarding what will happen to the existing account when a researcher left his/her current institution.	It is unclear what will happen to the existing account and uploaded data when a researcher has left his/her current institution. There should be information to address this issue.	UT	MO	TIC

E.3 BSHARE

E.3.1 List of problems found on B2SHARE

No	Problem	Description	UEM	Severity	Status
1	Sorting features are not working.	All sorting features are not working. Test users said that these problems make system the system looks unreliable.	FI, UT	CA	TIC

2	Finding similar records always display empty result.	When button to find similar records has been clicked, the system always displays empty result.	FI	MA	TIC
3	Dataset access restriction does not work.	Dataset access restriction is not functioning since the uploaded dataset can be downloaded by both registered and unregistered users regardless whether the selected Open Access option is ON or OFF.	FI	MA	FIX
4	Rates and comments do not appear on the dataset's description page.	Rates and comments given by other users do not appear on the dataset's description page.	FI	MA	TIC
5	Cloud connection does not work.	There are three options for cloud connection, whenever one of the options is clicked, the system always shows "Unexpected error".	FI	MI	TIC
6	Error appears when adding a dataset into the personal basket.	Internal server error always appears when a user adds one or more datasets into his/her personal basket. However, the datasets were added successfully into the personal basket.	FI	MI	REM
7	Error appears when copying or moving a dataset from different personal basket.	Internal server error always appears when a user copies or moves one or more datasets from different personal baskets. However, the datasets were copied and moved successfully into other personal baskets.	FI	MI	REM

8	Error appears when adding a note into a dataset in the personal basket.	Internal server error always appears when a user adds a note into a dataset in his/her personal basket.	FI	MI	REM
9	The button for alert WIP does not work.	The button for alert WIP is not functioning in the personal page. The system does not display anything when the button is clicked.	FI	MI	REM
10	Login fields are still exist after users have successfully logged in.	Users are not sure whether they have logged in or not since login fields are still appear on the interface.	FI, UT	CA	TIC
11	Voting (thumb-up or thumb-down) is not working.	Registered users are supposed to be able to vote on given comments on uploaded datasets. However, this feature is not working at the moment.	UT	MI	TIC
12	Problem when uploading a dataset.	After the user has filled the metadata and clicked the deposit button, the system showed please wait but there is nothing happened. After reloading the page, the user found that the uploading was unsuccessful. However, this problem occurred to one test user only.	UT	MI	TIC
13	Hide or remove button for adding tags on the description page of uploaded dataset.	Add tags button should not appear on the description page of dataset uploaded by other users. The button for adding tags should appear only for the dataset's uploader and administrators.	FI	MA	TIC

E.3.2 List of proposed suggestions for B2SHARE

No	Suggestion	Description	UEM	Priority	Status
1	Change the format for selecting collection to filter search results.	The current system changes the color and make the text bold when user clicked the collection. It would be better to use a checklist to filter search results according to the collection.	FI	MO	TIC
2	Sign out button should be available in every page.	Currently, it is difficult to sign out from the system since the sign out button is available in personal page only. The sign out button should be easy to find.	FI, UT	UR	TIC
3	Change the position of username and password forms.	The position of username and password forms should be in the centre.	FI	MO	IMP
4	Provide indexing or drop-down feature for user guides.	Currently, there are many information in the user guides. It would be better if the user guides provide indexing or drop-down feature. The complete information will be shown after the user clicked the title of the information.	FI	MO	TIC
5	Provide indexing or drop-down feature for FAQ.	Currently, there are much information in the FAQ. It would be better if the FAQ provides indexing or drop-down feature. The answer will be shown after the user clicked the specific question.	FI	MO	TIC

6	Display related disciplines to the selected domain.	The current system displays all disciplines regardless the selected domain. The system should display domain related disciplines only, unless the user has selected Generic as the domain.	FI	HI	TIC
7	Change the icon for add to basket.	The icon for add to basket should be modified to something more resemble the feature.	FI	MO	REJ
8	Modify the information regarding registration on the user guides.	Since the sign in with username and password will be disabled soon, it would be better to modify the information on the user guides. Therefore, the users will start registering their accounts by using B2ACCESS instead of normal registration.	FI	HI	TIC
9	Provide an option to download multiple files at one time.	Some datasets contain multiple files. Therefore, it would be useful for the users to have an option to download multiple les at one time.	FI, UT	UR	TIC
10	Make the alternative identifier clickable.	If the dataset has an alternative identifier, the alternative identifier should be clickable like the main identifier and link it where the dataset is stored.	FI	HI	TIC
11	Provide full data citation format.	In addition to citation for bibliography management, like BibTeX, EndNote, etc, every dataset should also have traditional formatted citation. Some researchers are still using the traditional formatted citation.	FI	MO	TIC

12	Display how many times the dataset has been cited.	Displaying how many times the dataset has been cited can encourage the researcher to share their data. In the same time, it can give a sense of reward to the researcher.	FI	LO	TIC
13	Display the search result that would match the query semantically.	Currently, the system only displays search results that would match the query syntactically. The system should also display search results that would match the query semantically.	FI	HI	TIC
14	Display the most popular datasets on the homepage of B2SHARE.	Currently, the system only displays newest uploaded datasets on the homepage of B2SHARE. The system should also display most popular datasets on the homepage of B2SHARE. It can give a sense of reward to the researcher.	FI	MO	TIC
15	Provide profile to show user contribution.	Show information to the user about how many datasets that he/she has uploaded, or how many times his/her deposits have been downloaded. It can give a sense of reward to the researcher.	FI, UT	MO	TIC
16	Add search feature for selecting disciplines.	Selecting disciplines are difficult. The list is too long and there are few disciplines that cannot be read since it is too long. There should be a search function to find suitable disciplines.	UT	UR	TIC

17	Provide navigation menu on B2SHARE.	It is easy to get lost while navigating in B2SHARE. There should be a way to inform users about their location and how to move from one page to other pages.	UT	UR	TIC
18	Move the button for start uploading to the below or next to the list of uploaded files.	Test users were not sure whether B2SHARE is automatically uploading files or not. Button for start uploading should be located below or next to list of uploaded files.	UT	HI	TIC
19	Provide short description about what the domain is.	Most test users assumed domain means research domains or disciplines. There should be a short explanation to describe what the domain is.	UT	UR	TIC
20	Provide a calendar-like option for selecting the date of embargo.	There should be a calendar-like option when users want to select the embargo. It should not be in a free-form field.	UT	HI	TIC
21	New uploaded dataset should be available and can be searched immediately.	Currently, it took five to ten minutes for a recently uploaded dataset to be publicly available. The new uploaded dataset should be available and can be searched as soon as it has been uploaded.	UT	HI	TIC

22	Provide short description how to fill the owner organization.	Test users were not sure regarding the meaning of owner organization when they have selected Aalto as the domain. It is unclear whether they should put their department or school or simply put Aalto University. There should be information to address this issue when users have selected Aalto as the domain of their datasets.	UT	UR	TIC
23	Simplify the process to give a rating.	Users should not need to select the star again after they clicked it first time. The stars should not look different after being clicked.	UT	MO	TIC
24	Provide short description for each terminology.	There are some terms that are not familiar for the test users, such as domain, deposit, embargo, and license. There should be a short description for each term.	UT	UR	TIC
25	Provide an overview of all uploaded datasets.	In addition to search function, there should be one option to explore or check what kind of datasets are available on B2SHARE. An overview would be a useful feature.	UT	MO	TIC
26	Change deposit into upload.	It would be better to use upload, instead of deposit since upload has clearer meaning.	UT	MO	TIC
27	Remove one of the language fields.	It is not clear why there are two fields for language when uploading dataset. If both are referring to the same thing, B2SHARE should remove one of them.	UT	MO	TIC

28	Possibility to upload folder.	Some researchers have research data that have been structured. Possibility to upload the whole folder would be useful.	UT	LO	TIC
29	Mention the limitation of B2SHARE.	Currently, users could not upload a file larger than 3 GB. B2SHARE should mention about this kind of limitation when users are uploading their datasets.	UT	UR	TIC
30	Change the download button to request dataset when the dataset has restricted access on the search results.	On the search results, the system will show an empty list when a user clicked the download button if the dataset has restricted access. The download button should be changed to request button.	UT	HI	TIC
31	Provide one way to fill the keywords.	Currently, there are two ways to ll the keywords: by putting comma or by using separate lines. However, this is confusing since it is not clear whether users should put keywords in one line by using comma or using separate lines. Having only one way to fill the keywords would be easier.	UT	MO	TIC
32	Provide an icon to show the access status of a dataset.	On the search results, B2SHARE should show whether the dataset is open or restricted. Users can only check the access option after they opened the description page of the dataset.	UT	HI	TIC

33	Possibility to add additional files into existing uploaded dataset.	Currently, users cannot upload additional files once the dataset has been uploaded. Possibility to add additional files into uploaded dataset would be useful since some files may need to be revised.	UT	MO	TIC
34	Provide information regarding how secure and safe B2SHARE is.	Security and privacy are one of the main considerations for researchers. B2SHARE should provide some information regarding security and privacy of its service.	UT	HI	TIC

E.4 B2ACCESS

E.4.1 List of problems found on B2ACCESS

No	Problem	Description	UEM	Severity	Status
1	There is an error when logging in after creating an account.	An error appears when the user is trying to log in after successfully creating a new account on B2ACCESS.	FI, UT	CA	TIC
2	B2ACCESS's terms of use and data privacy statement is not clickable.	Currently, the system displays terms of use and data privacy statement that should be accepted by the users when they are creating an account. However, the terms of use and data privacy statement are not clickable and cannot be read.	UT	MA	FIX

E.4.2 List of proposed suggestions for B2ACCESS

No	Suggestion	Description	UEM	Priority	Status
1	Provide meaningful messages on the system.	Test users were not sure what is happening when they are registering and logging in to B2SHARE via B2ACCESS. Information shown by the system to the users during these processes are not helpful.	UT	UR	TIC
2	Registration and login process should be shortened.	Currently, there are many steps that need to be taken in order to register and log in to B2ACCESS.	UT	UR	TIC
3	Include information about how to fix the error on the error messages.	The error messages shown on the system are not helpful. The error messages should also mention how users can prevent or solve the errors.	UT	UR	TIC

Appendix F

Results from post-task questionnaires

F.1 Results from System Usability Scale (SUS)

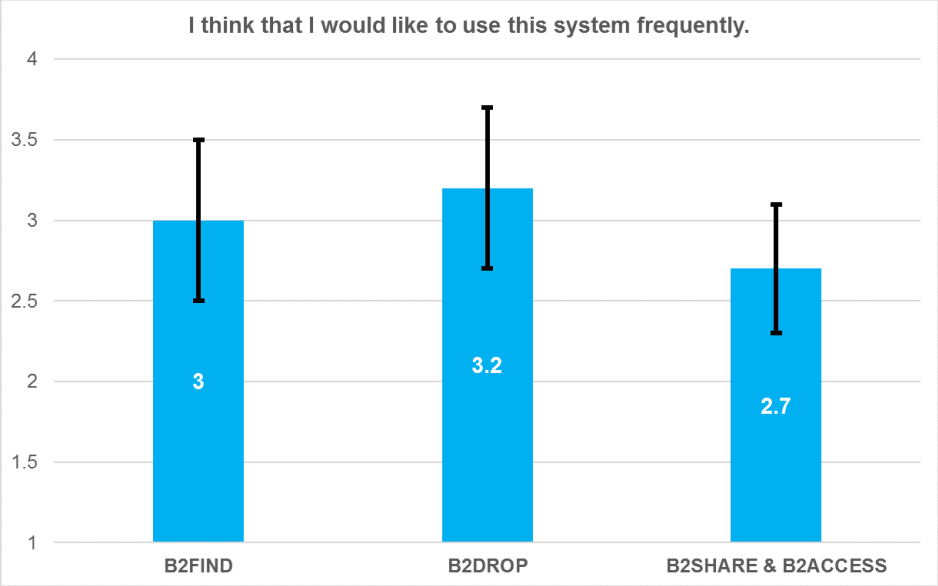

Figure F.1: The mean scores of test users' responses on statement no. 1 in SUS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

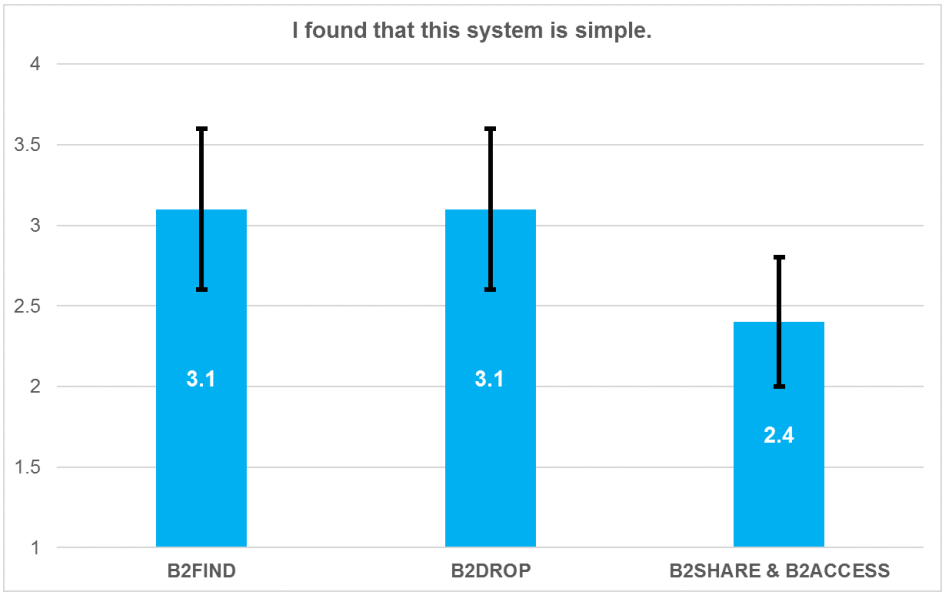

Figure F.2: The mean scores of test users' responses on statement no. 2 in SUS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

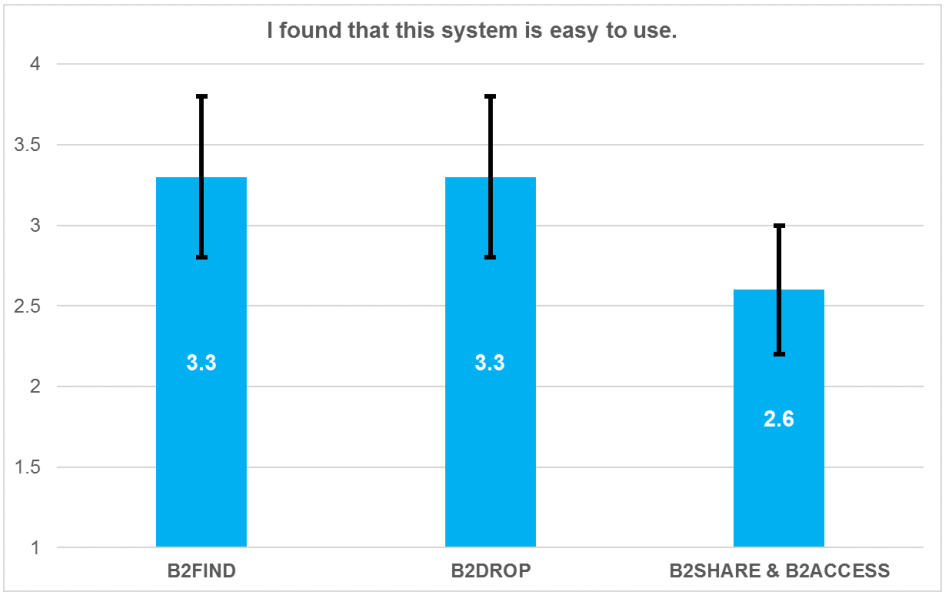

Figure F.3: The mean scores of test users' responses on statement no. 3 in SUS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

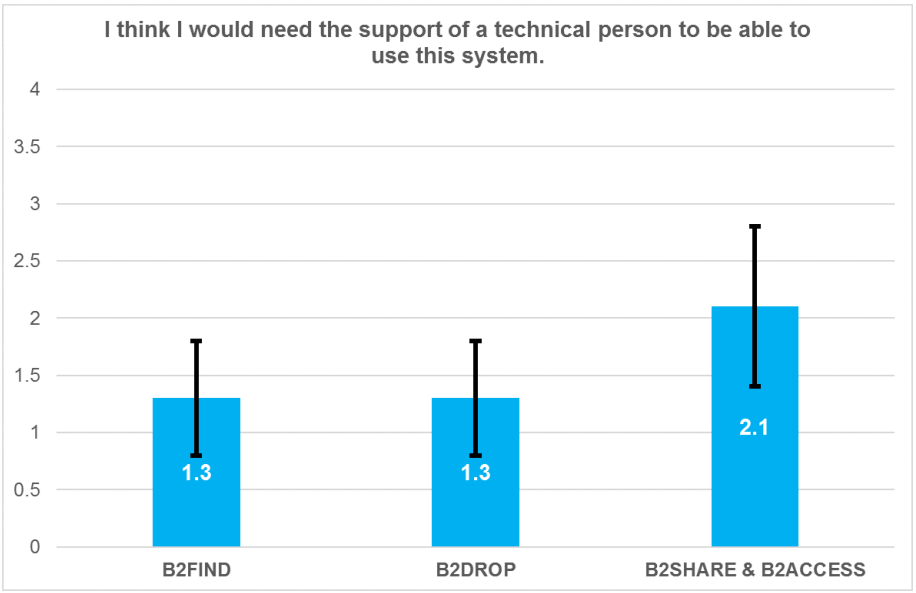

Figure F.4: The mean scores of test users' responses on statement no. 4 in SUS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

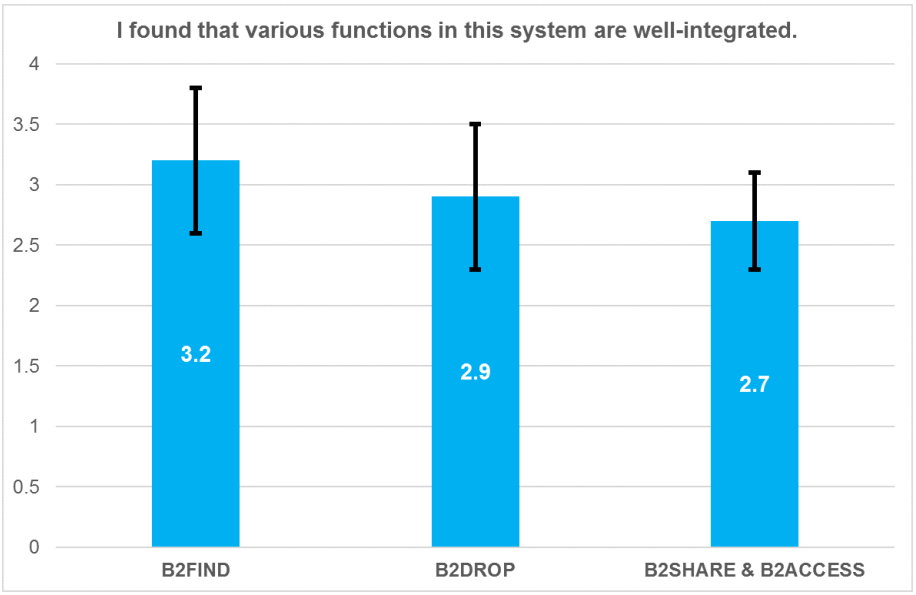

Figure F.5: The mean scores of test users' responses on statement no. 5 in SUS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

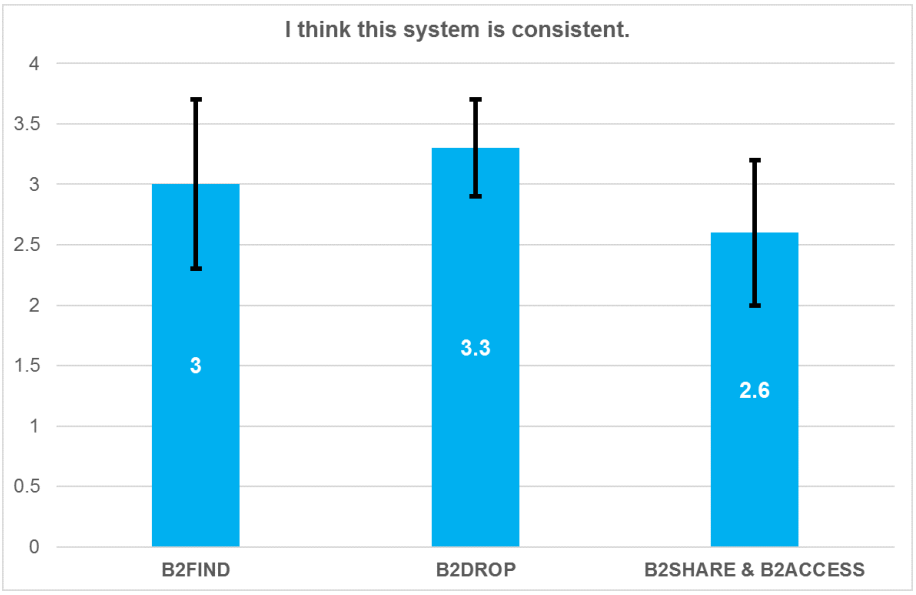

Figure F.6: The mean scores of test users' responses on statement no. 6 in SUS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

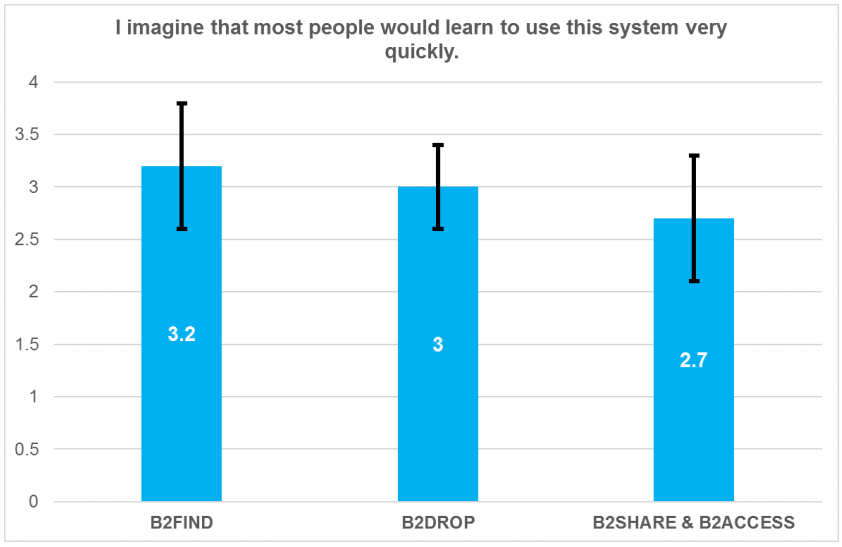

Figure F.7: The mean scores of test users' responses on statement no. 7 in SUS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

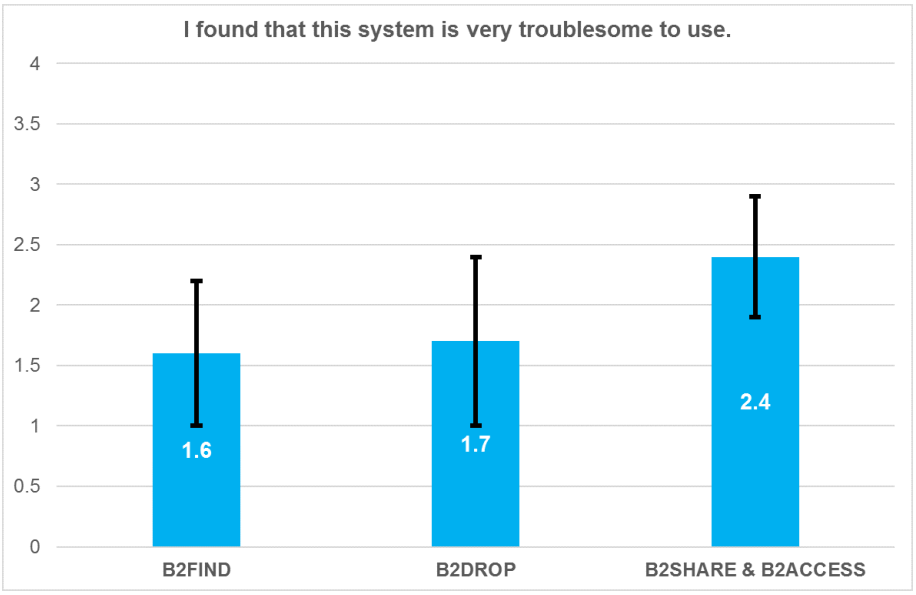

Figure F.8: The mean scores of test users' responses on statement no. 8 in SUS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

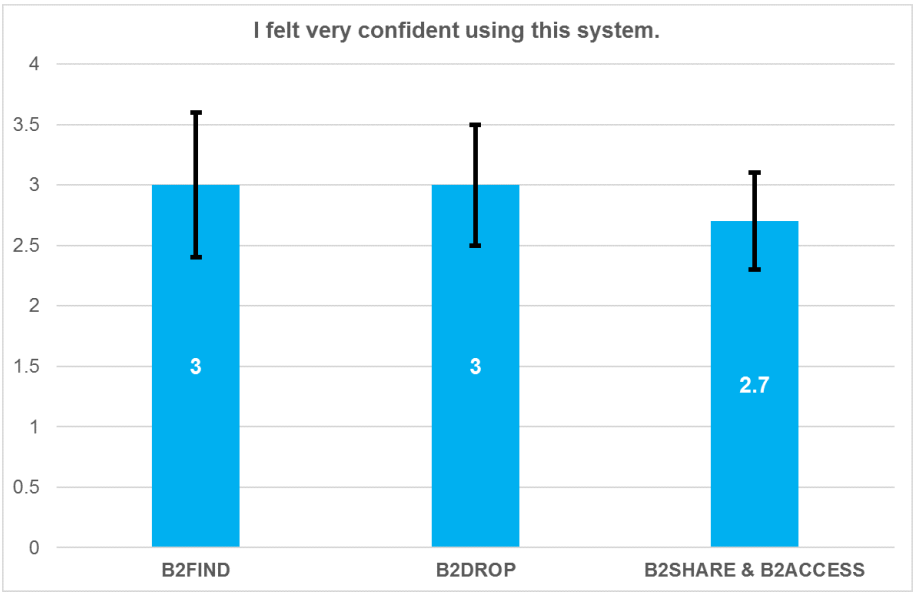

Figure F.9: The mean scores of test users' responses on statement no. 9 in SUS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

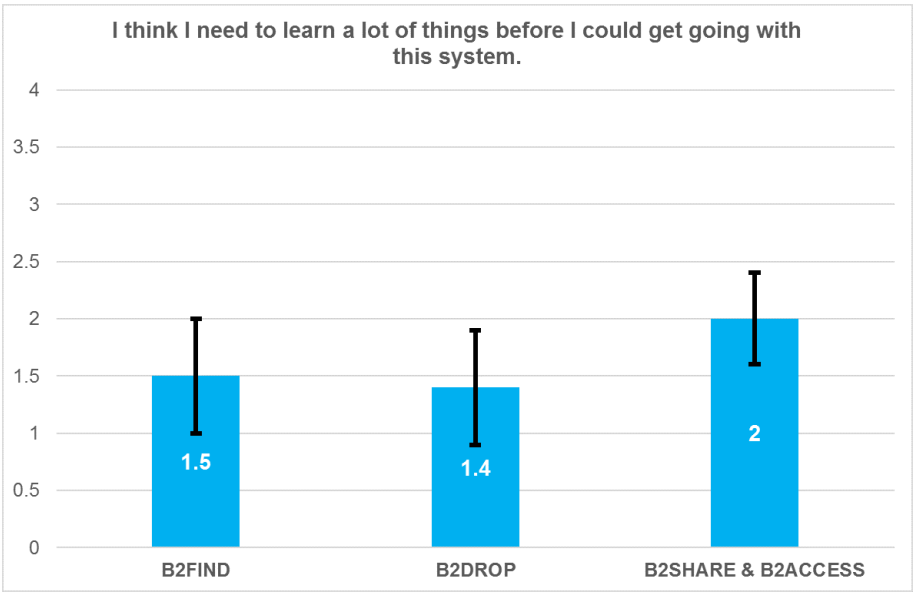

Figure F.10: The mean scores of test users' responses on statement no. 10 in SUS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

F.2 Results from Computer System Usability Questionnaire (CSUQ)

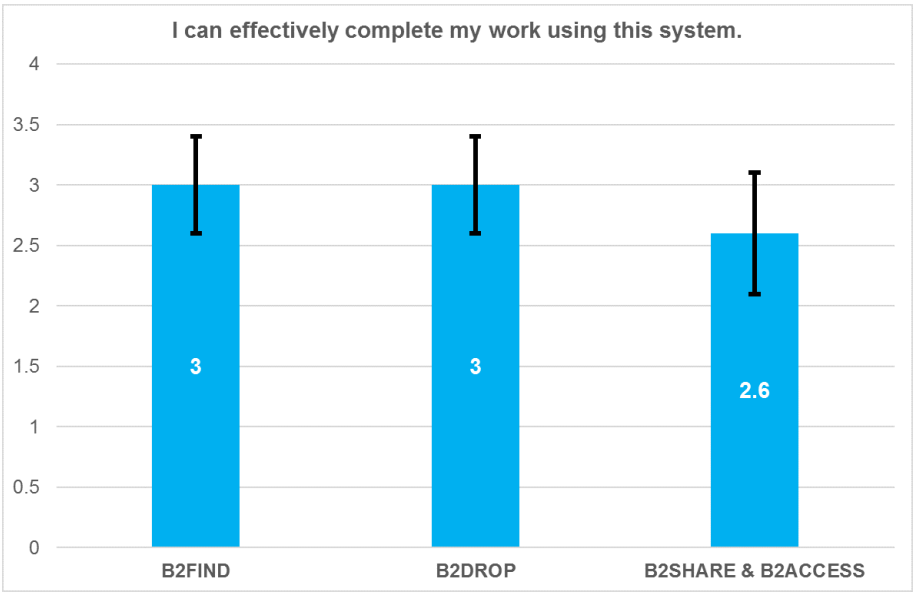

Figure F.11: The mean scores of test users' responses on statement no. 1 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

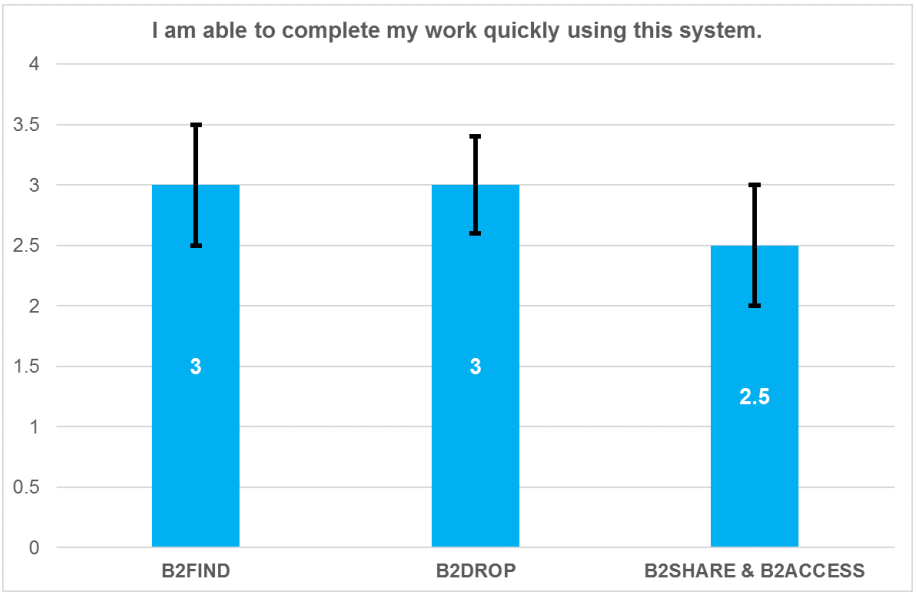

Figure F.12: The mean scores of test users' responses on statement no. 2 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

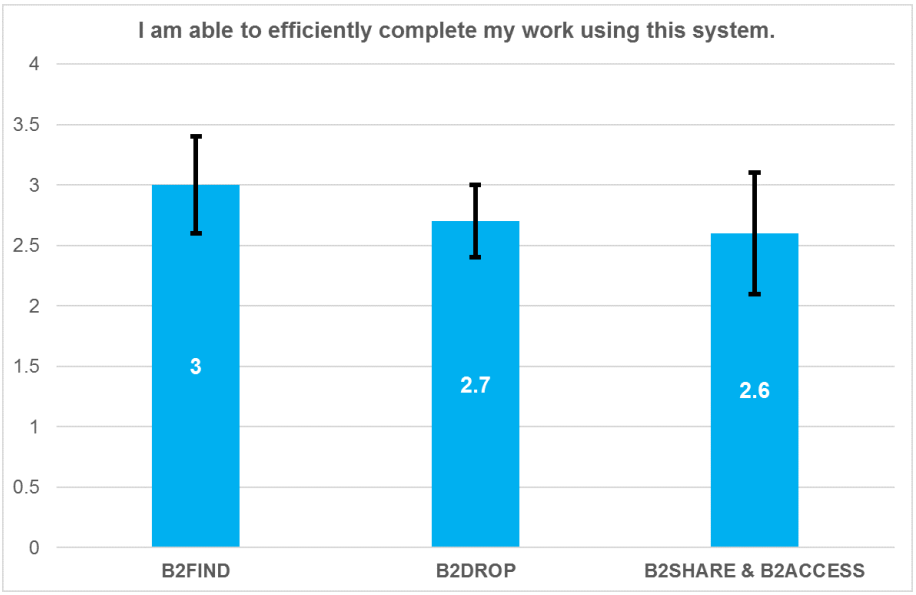

Figure F.13: The mean scores of test users' responses on statement no. 3 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

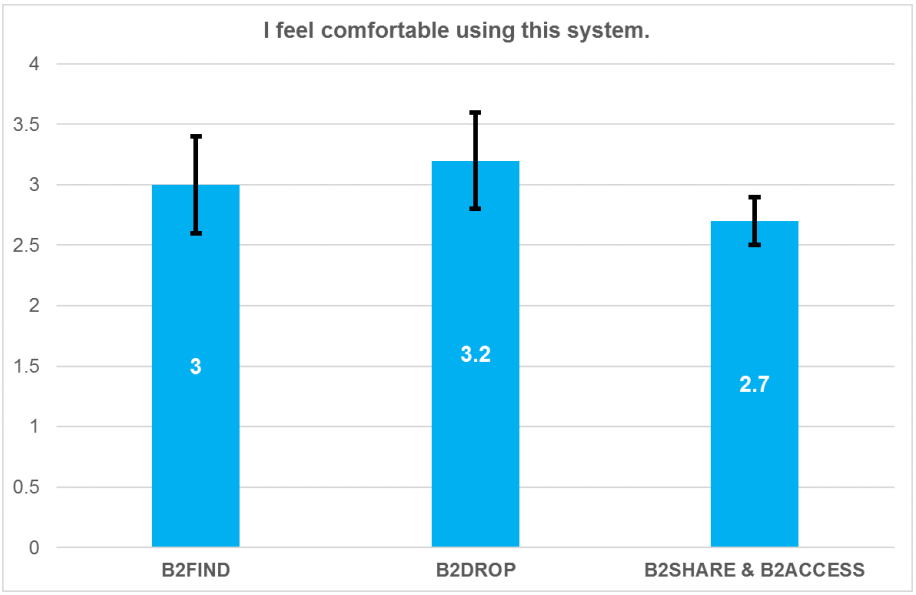

Figure F.14: The mean scores of test users' responses on statement no. 4 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

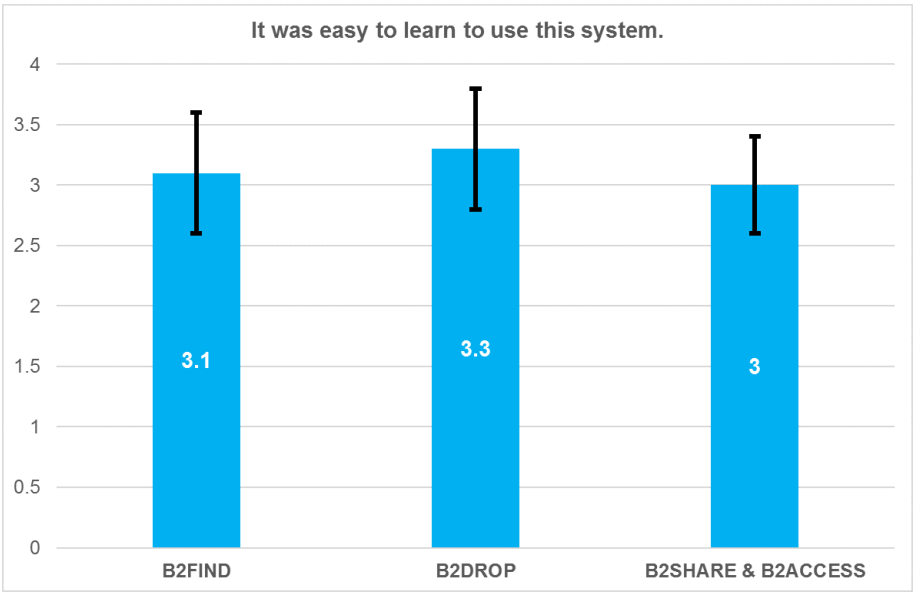

Figure F.15: The mean scores of test users' responses on statement no. 5 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

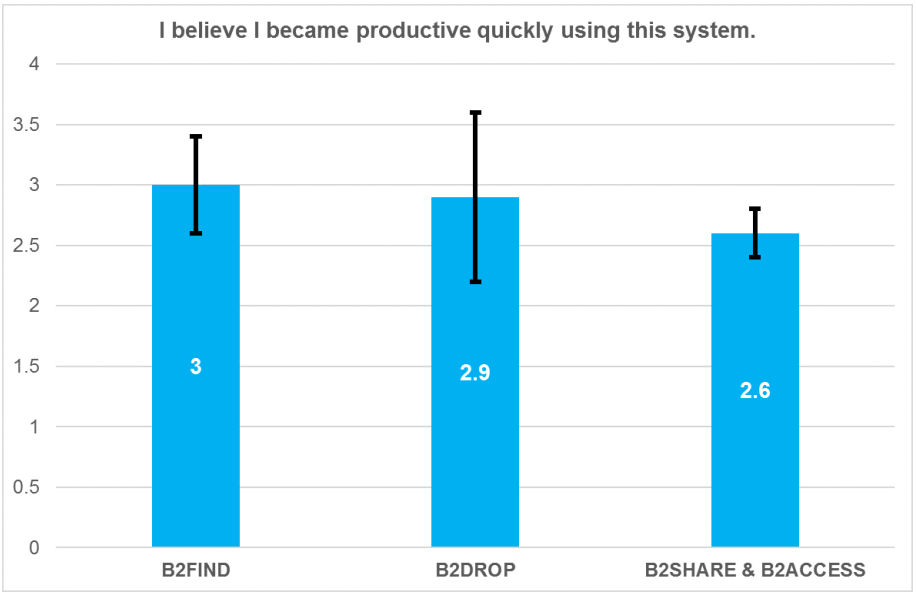

Figure F.16: The mean scores of test users' responses on statement no. 6 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

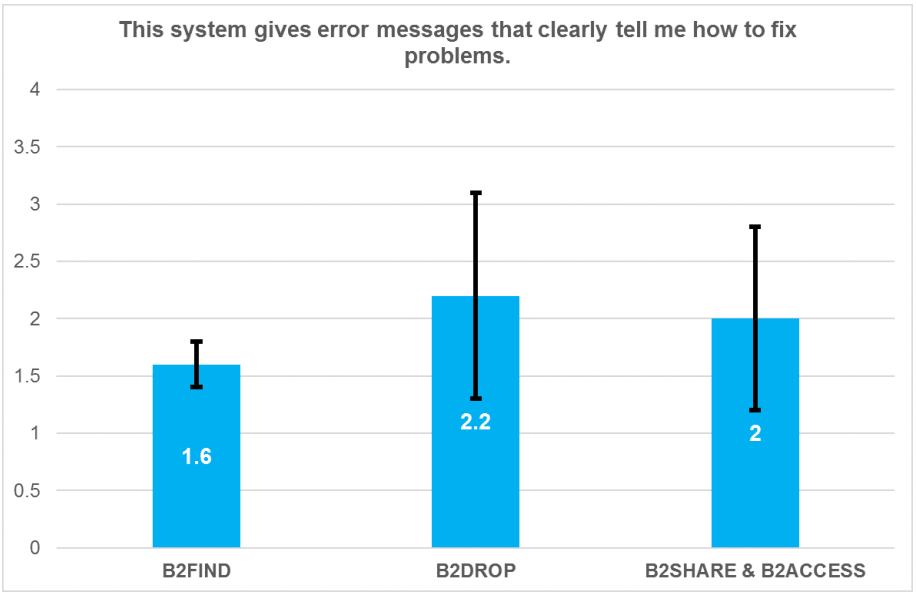

Figure F.17: The mean scores of test users' responses on statement no. 7 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

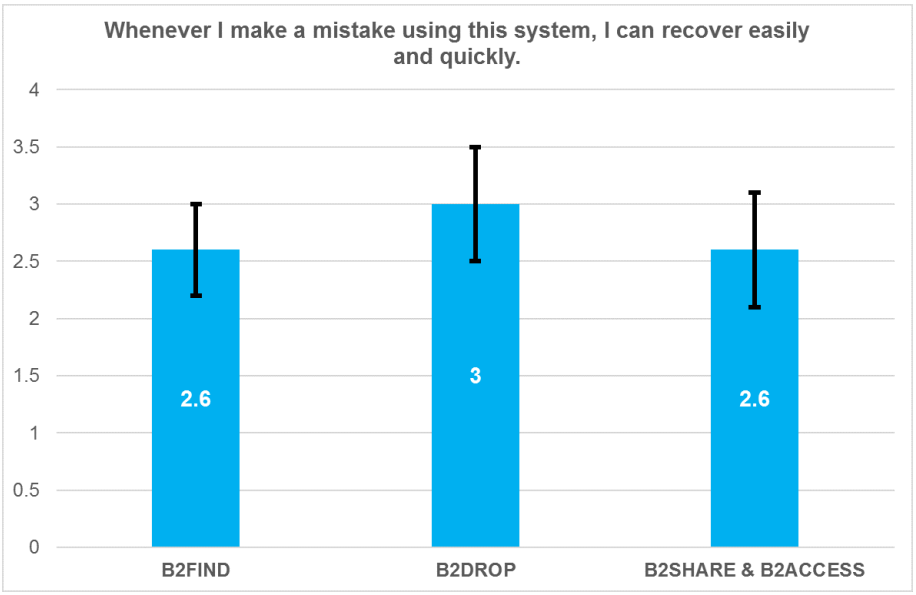

Figure F.18: The mean scores of test users' responses on statement no. 8 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

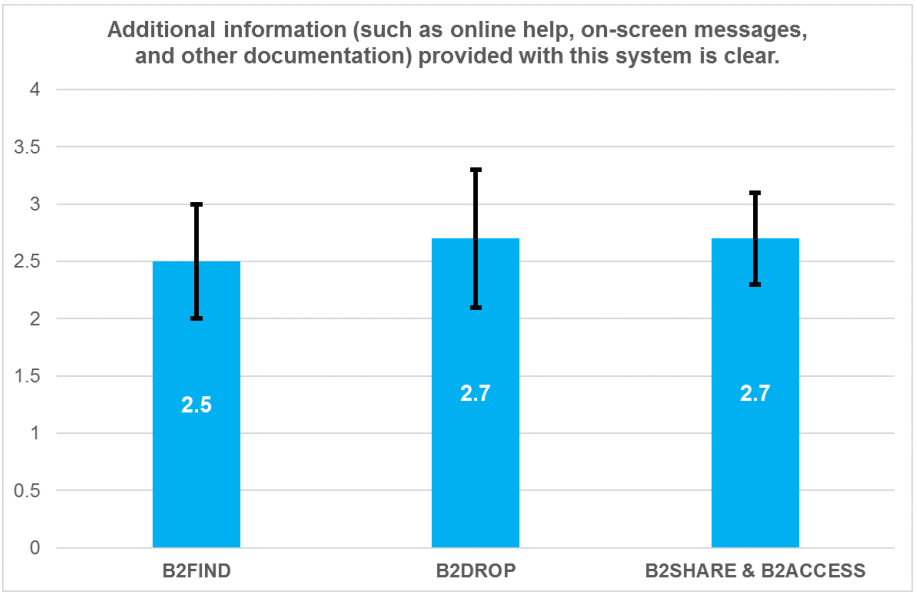

Figure F.19: The mean scores of test users' responses on statement no. 9 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

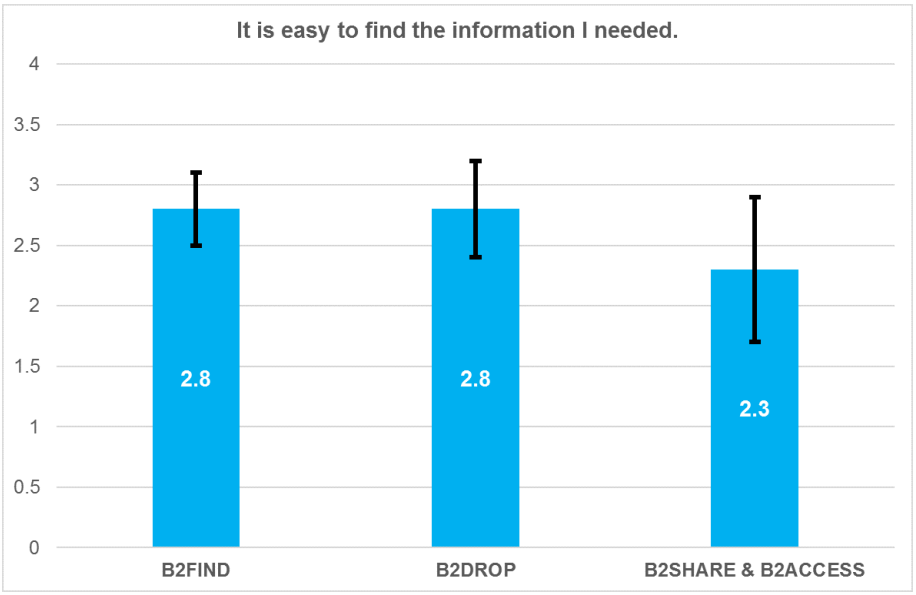

Figure F.20: The mean scores of test users' responses on statement no. 10 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

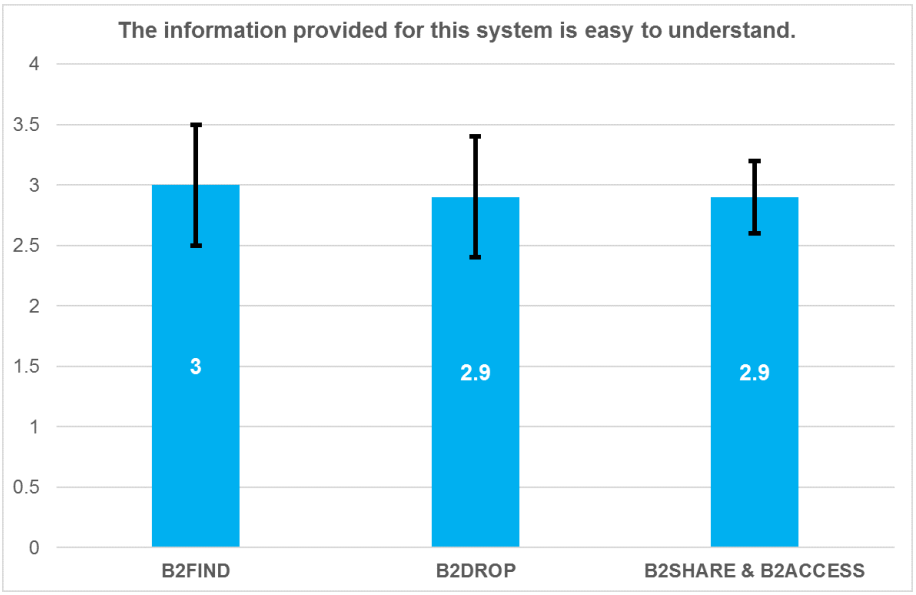

Figure F.21: The mean scores of test users' responses on statement no. 11 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

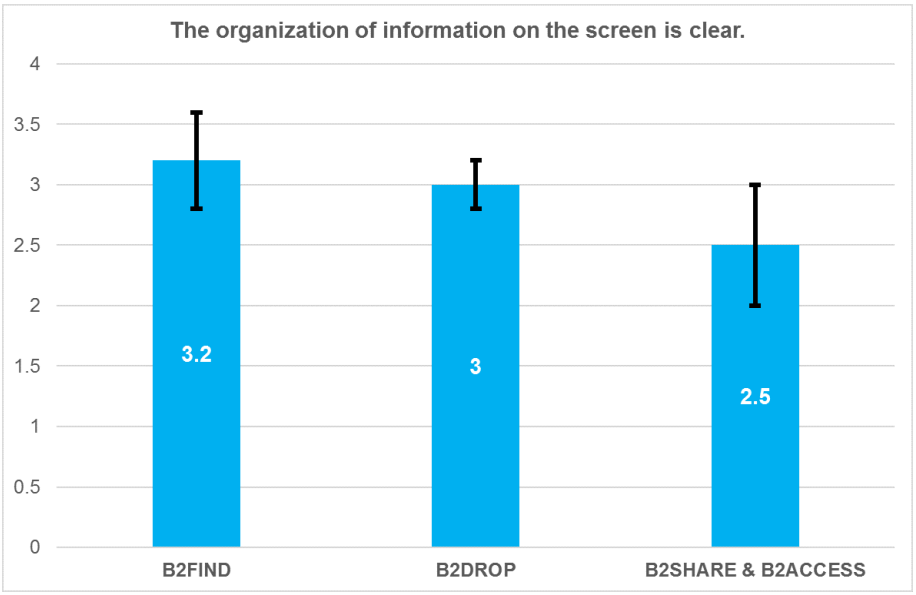

Figure F.22: The mean scores of test users' responses on statement no. 12 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

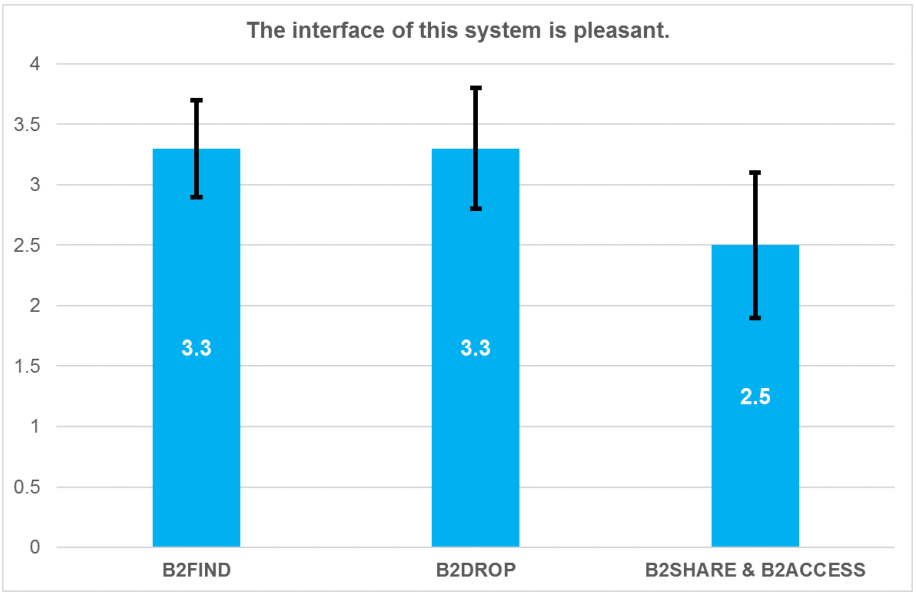

Figure F.23: The mean scores of test users' responses on statement no. 13 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

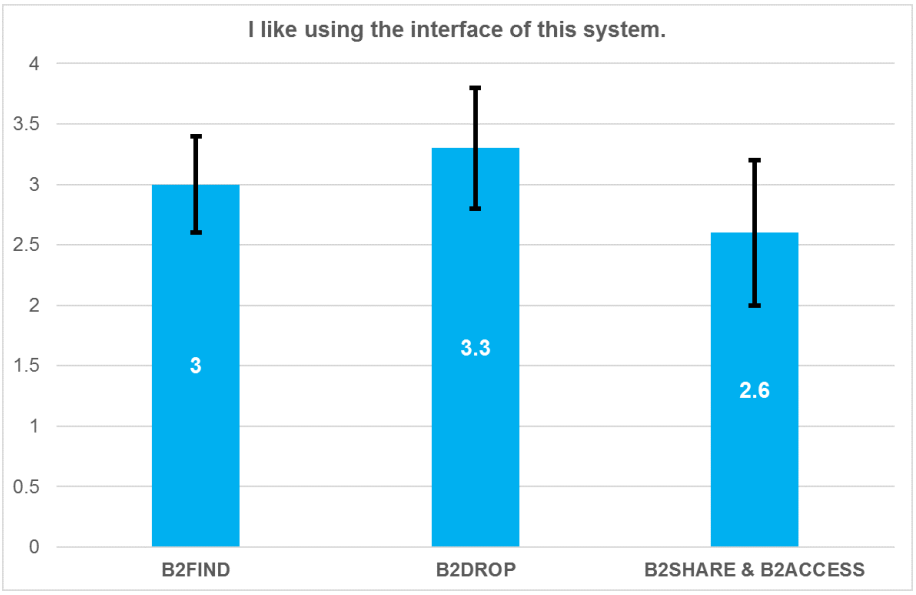

Figure F.24: The mean scores of test users' responses on statement no. 14 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

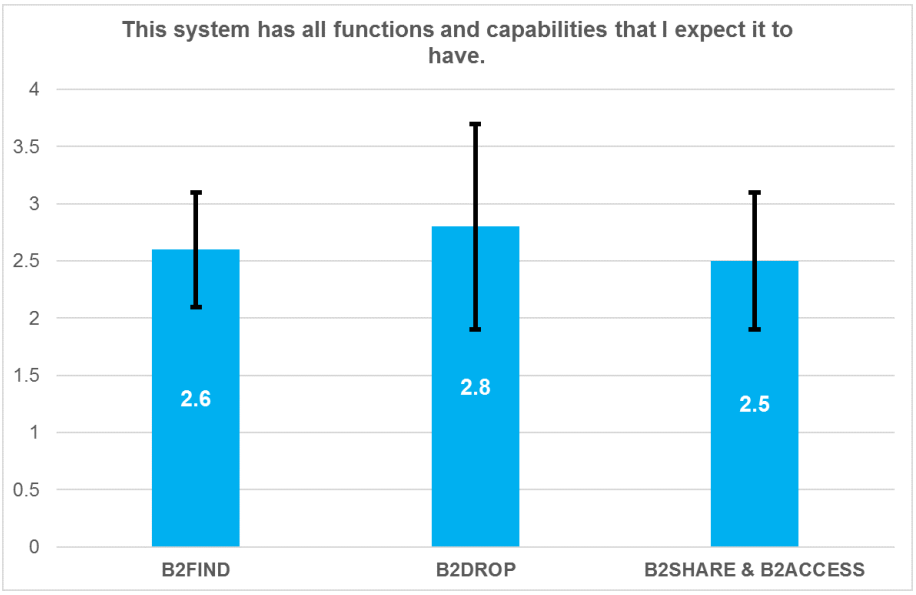

Figure F.25: The mean scores of test users' responses on statement no. 15 in CSUQ (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

F.3 Results from Questionnaire for User Interaction Satisfaction (QUIS)

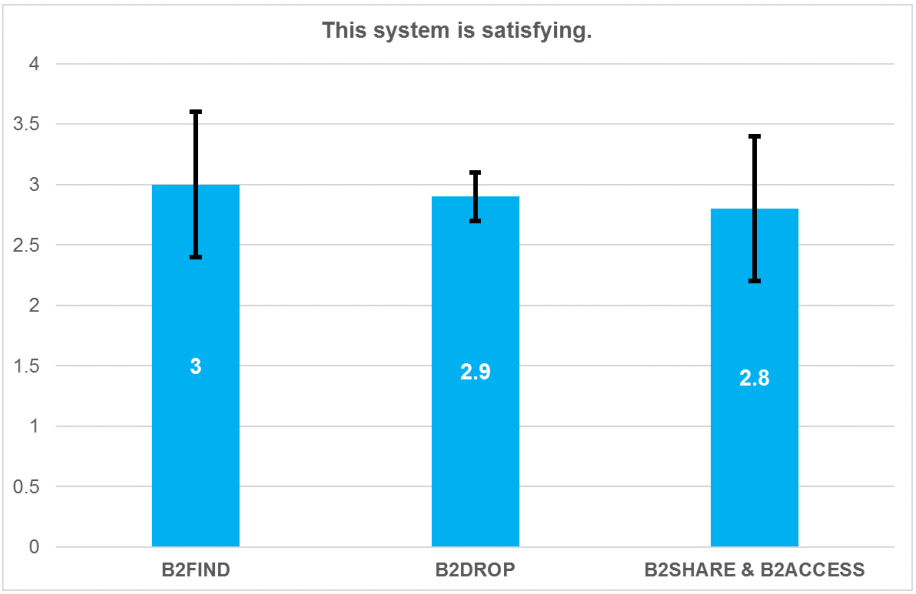

Figure F.26: The mean scores of test users' responses on statement no. 1 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

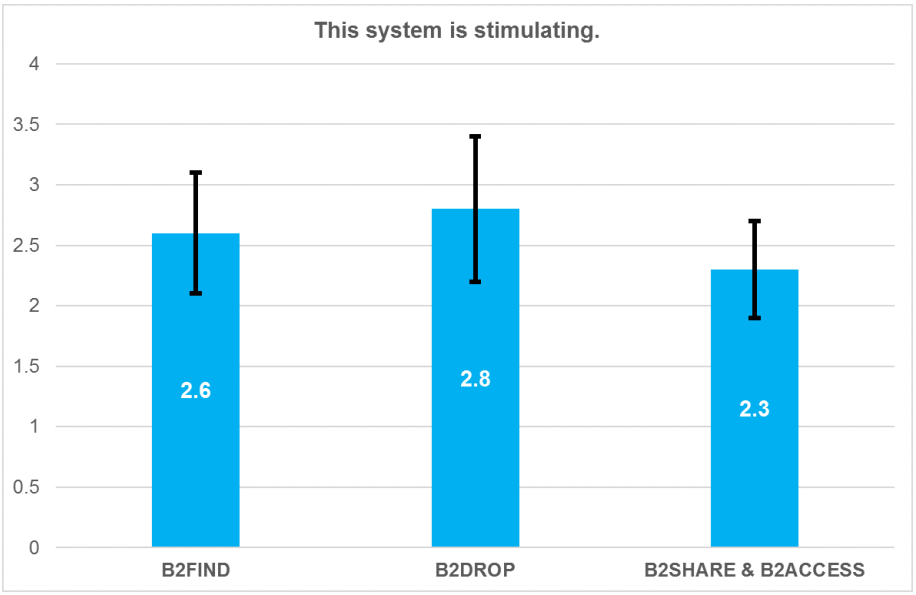

Figure F.27: The mean scores of test users' responses on statement no. 2 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

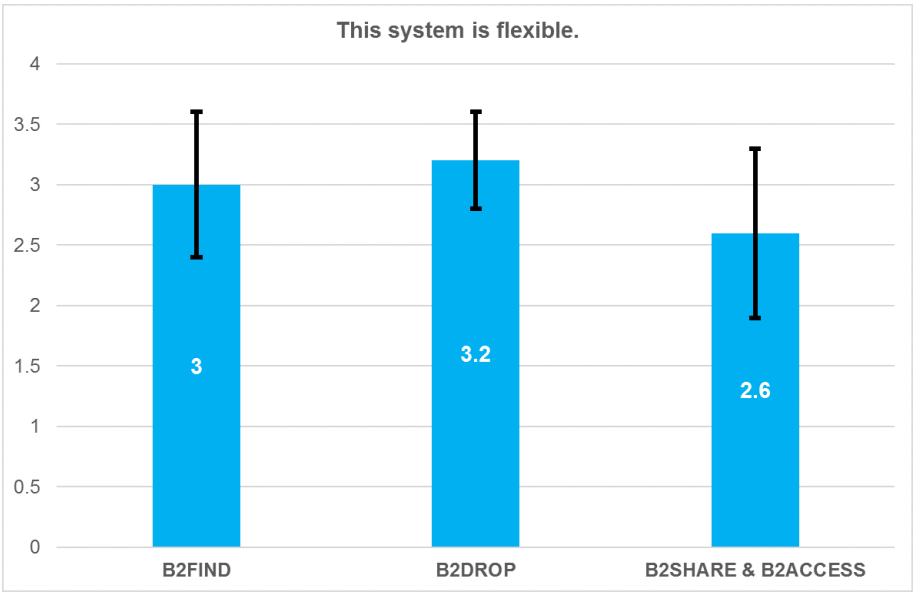

Figure F.28: The mean scores of test users' responses on statement no. 3 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

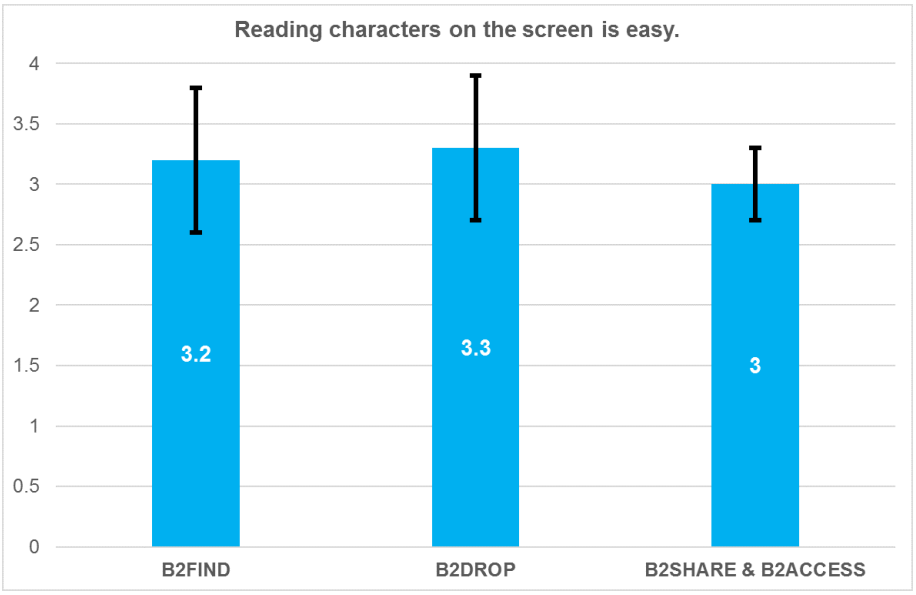

Figure F.29: The mean scores of test users' responses on statement no. 4 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

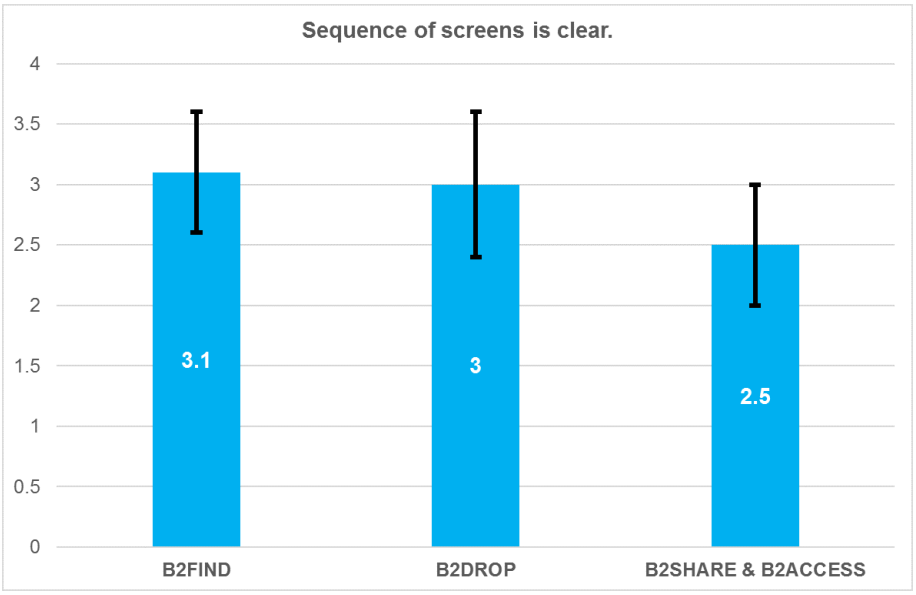

Figure F.30: The mean scores of test users' responses on statement no. 5 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

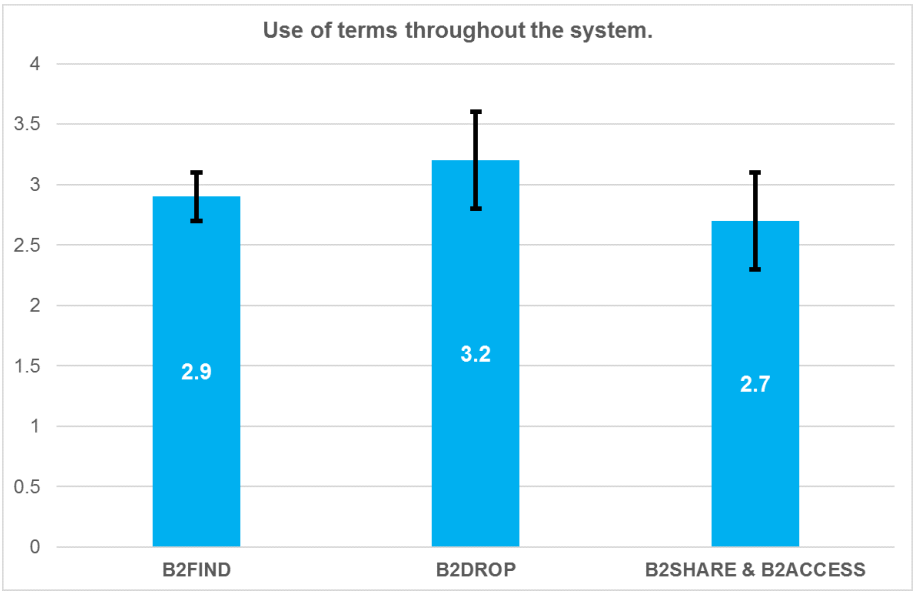

Figure F.31: The mean scores of test users' responses on statement no. 6 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

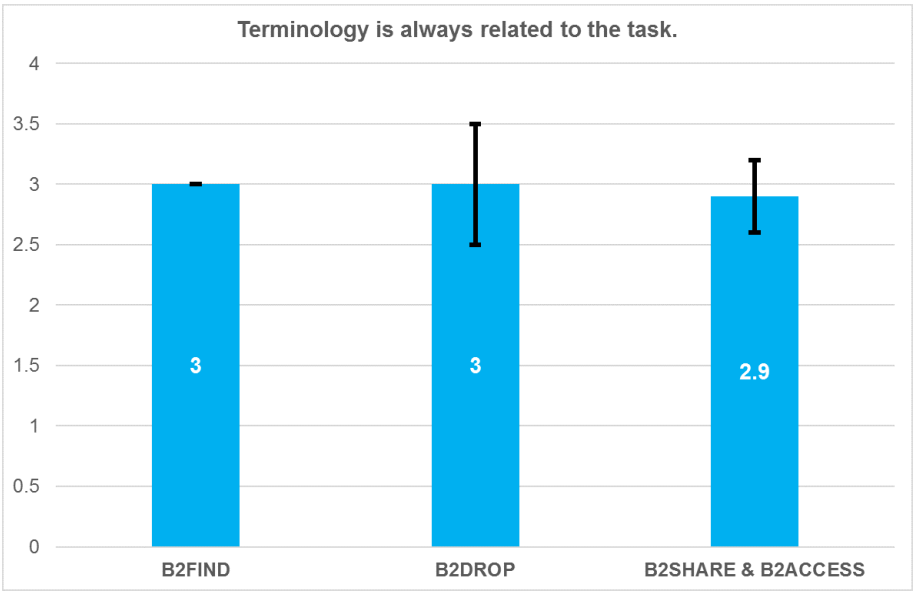

Figure F.32: The mean scores of test users' responses on statement no. 7 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

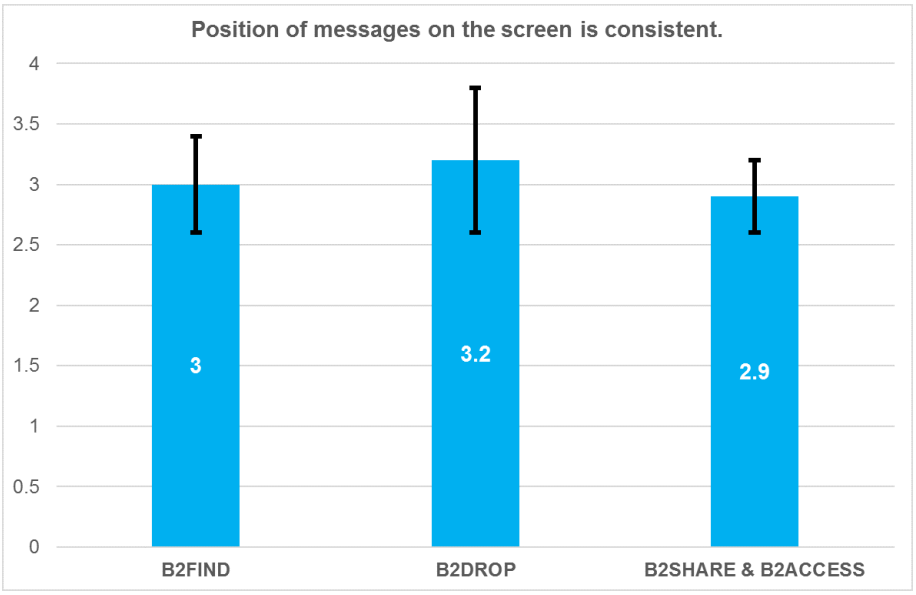

Figure F.33: The mean scores of test users' responses on statement no. 8 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

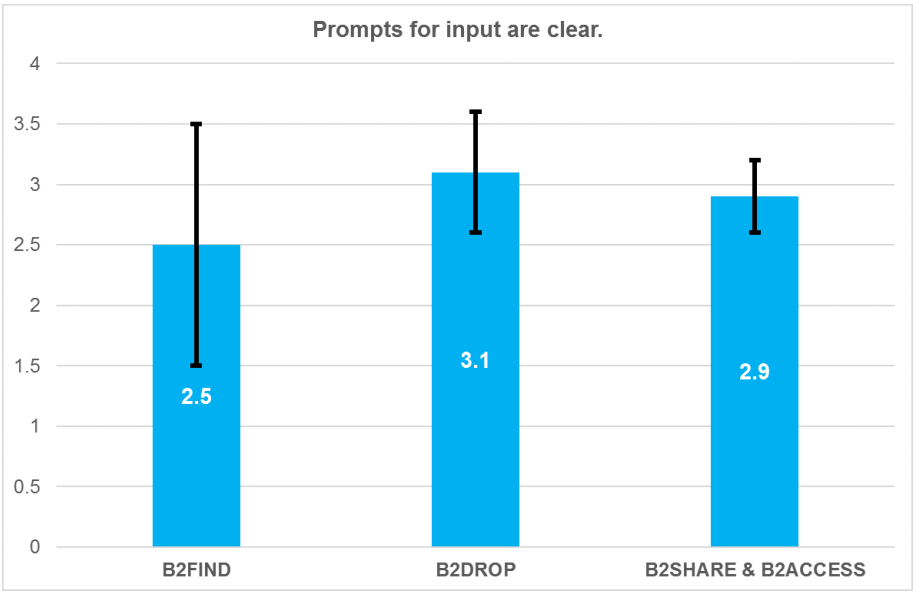

Figure F.34: The mean scores of test users' responses on statement no. 9 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

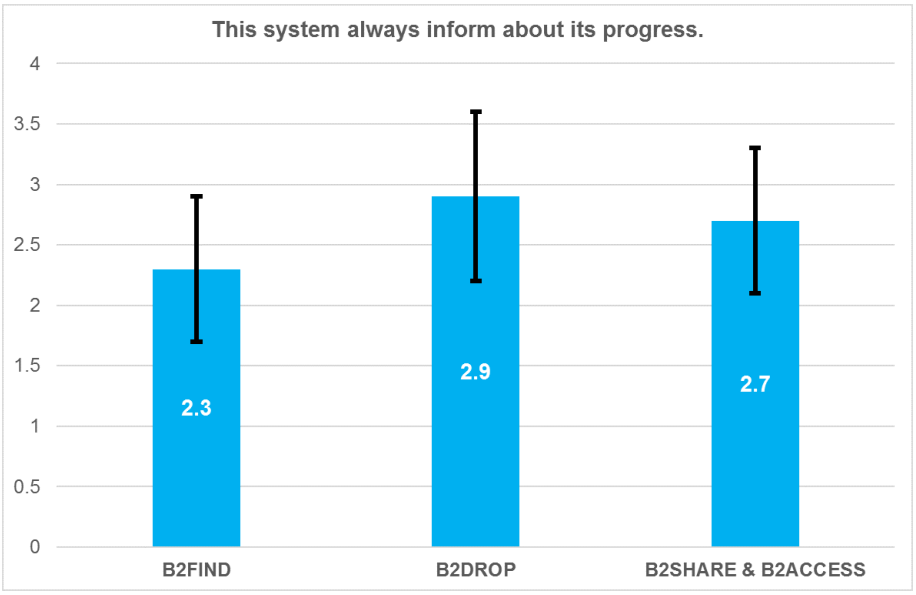

Figure F.35: The mean scores of test users' responses on statement no. 10 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

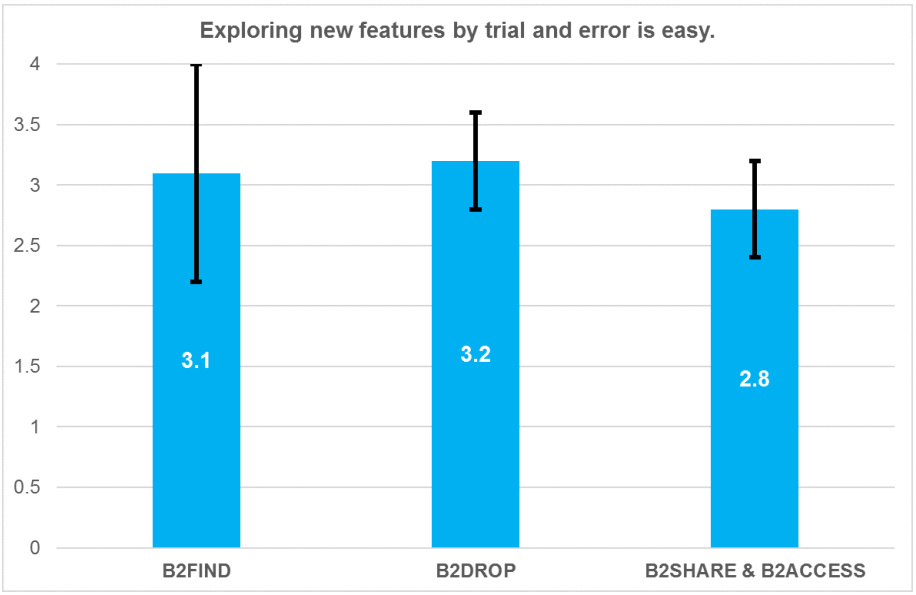

Figure F.36: The mean scores of test users' responses on statement no. 11 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

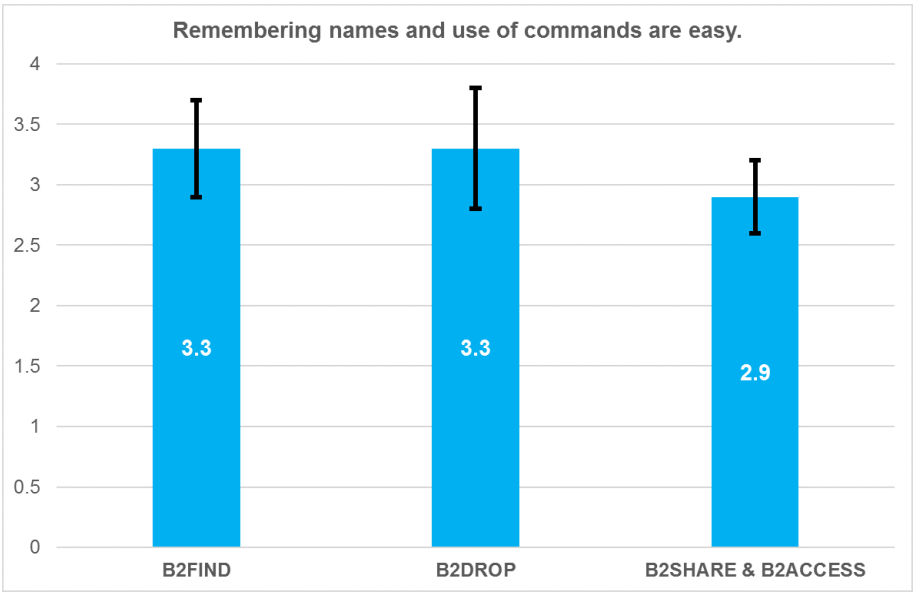

Figure F.37: The mean scores of test users' responses on statement no. 12 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

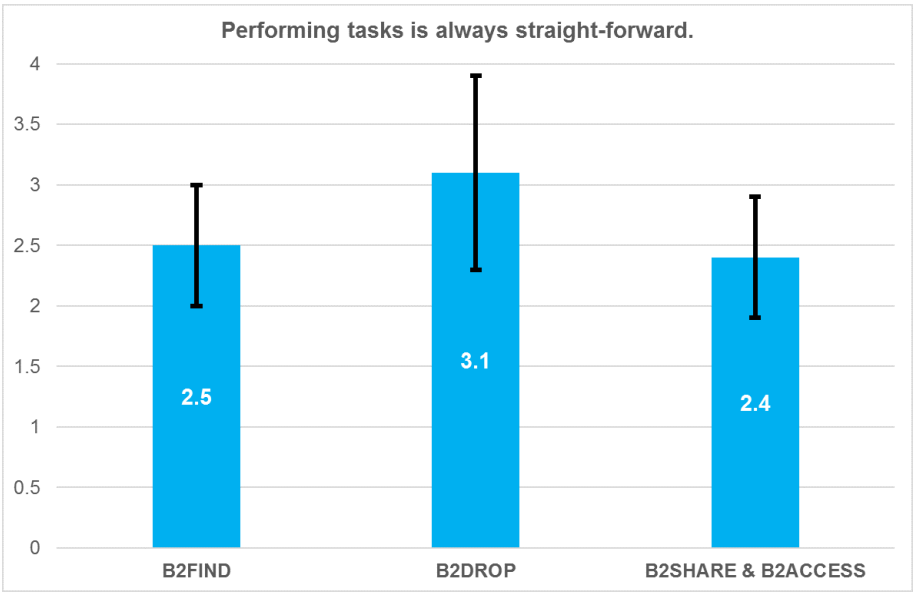

Figure F.38: The mean scores of test users' responses on statement no. 13 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

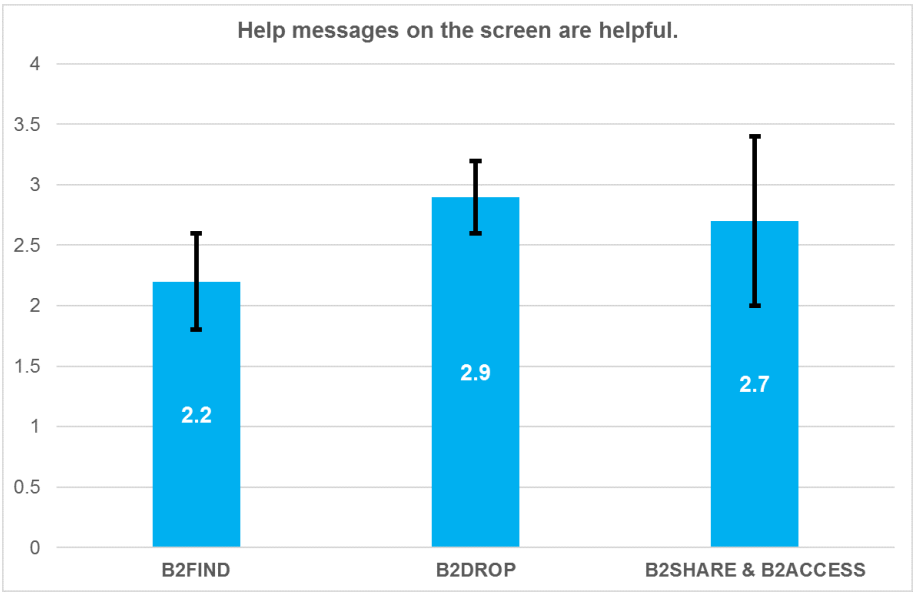

Figure F.39: The mean scores of test users' responses on statement no. 14 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

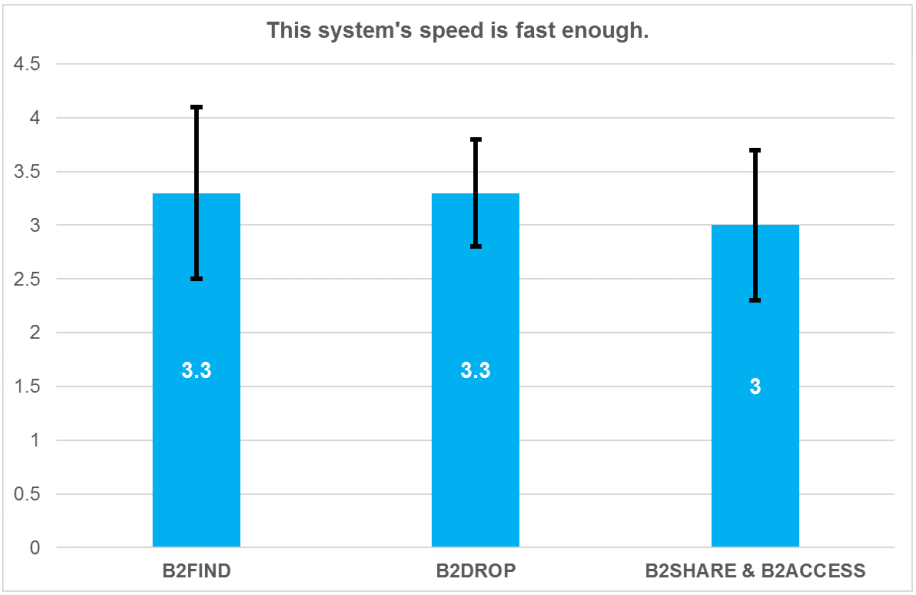

Figure F.40: The mean scores of test users' responses on statement no. 15 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

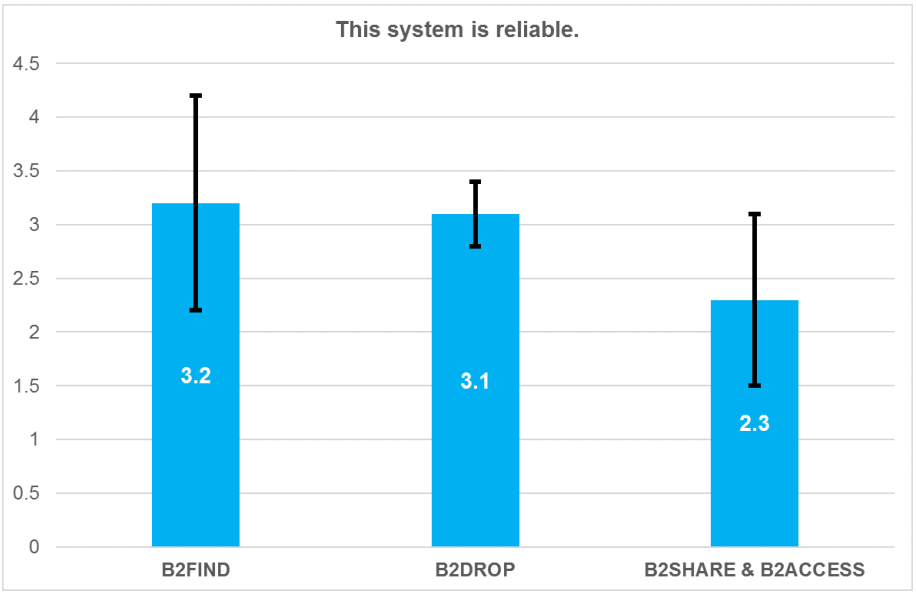

Figure F.41: The mean scores of test users' responses on statement no. 16 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

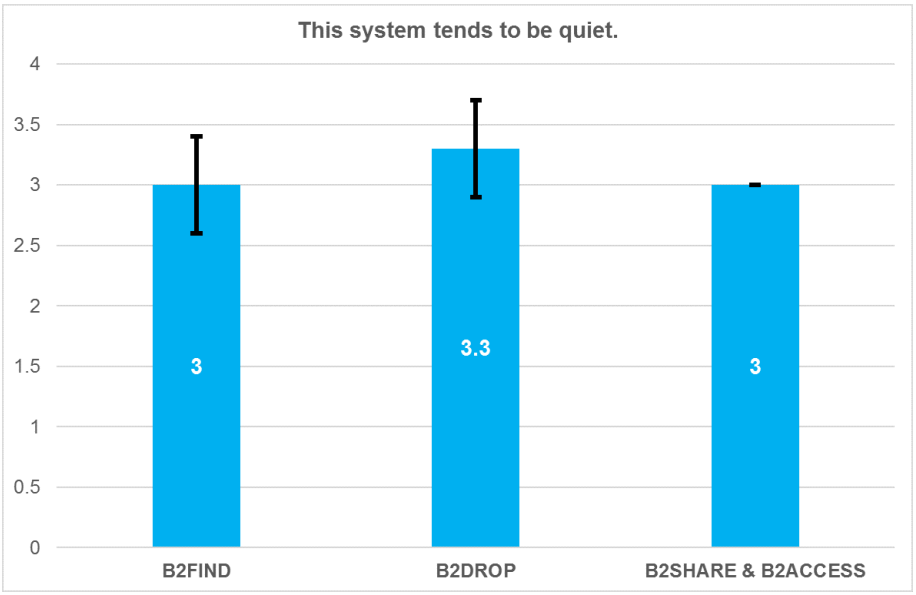

Figure F.42: The mean scores of test users' responses on statement no. 17 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4

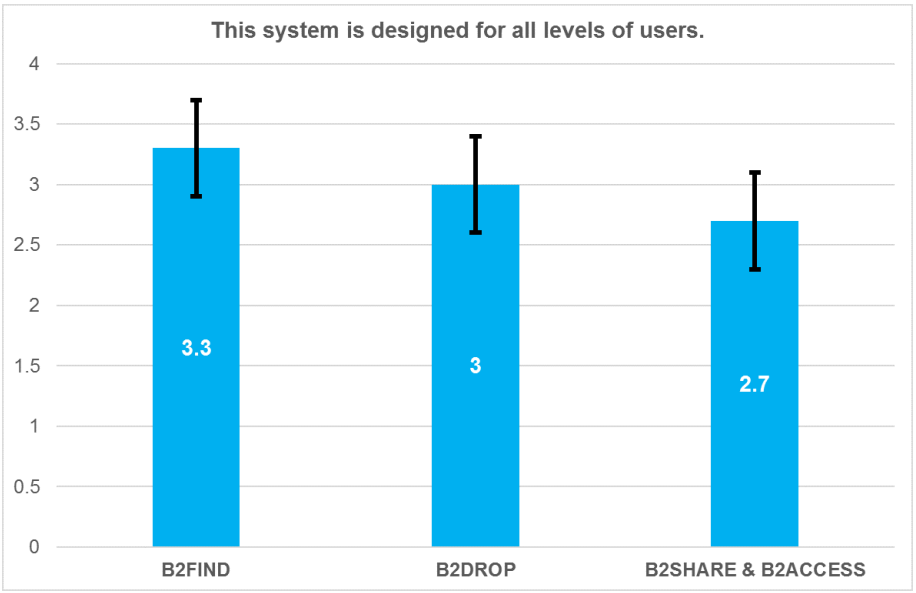

Figure F.43: The mean scores of test users' responses on statement no. 18 in QUIS (The black bars represent the standard deviation of the mean). Strongly disagree = 1, disagree = 2, agree = 3, and strongly agree = 4